UC San Diego
JACOBS SCHOOL OF ENGINEERING
Electrical and Computer Engineering

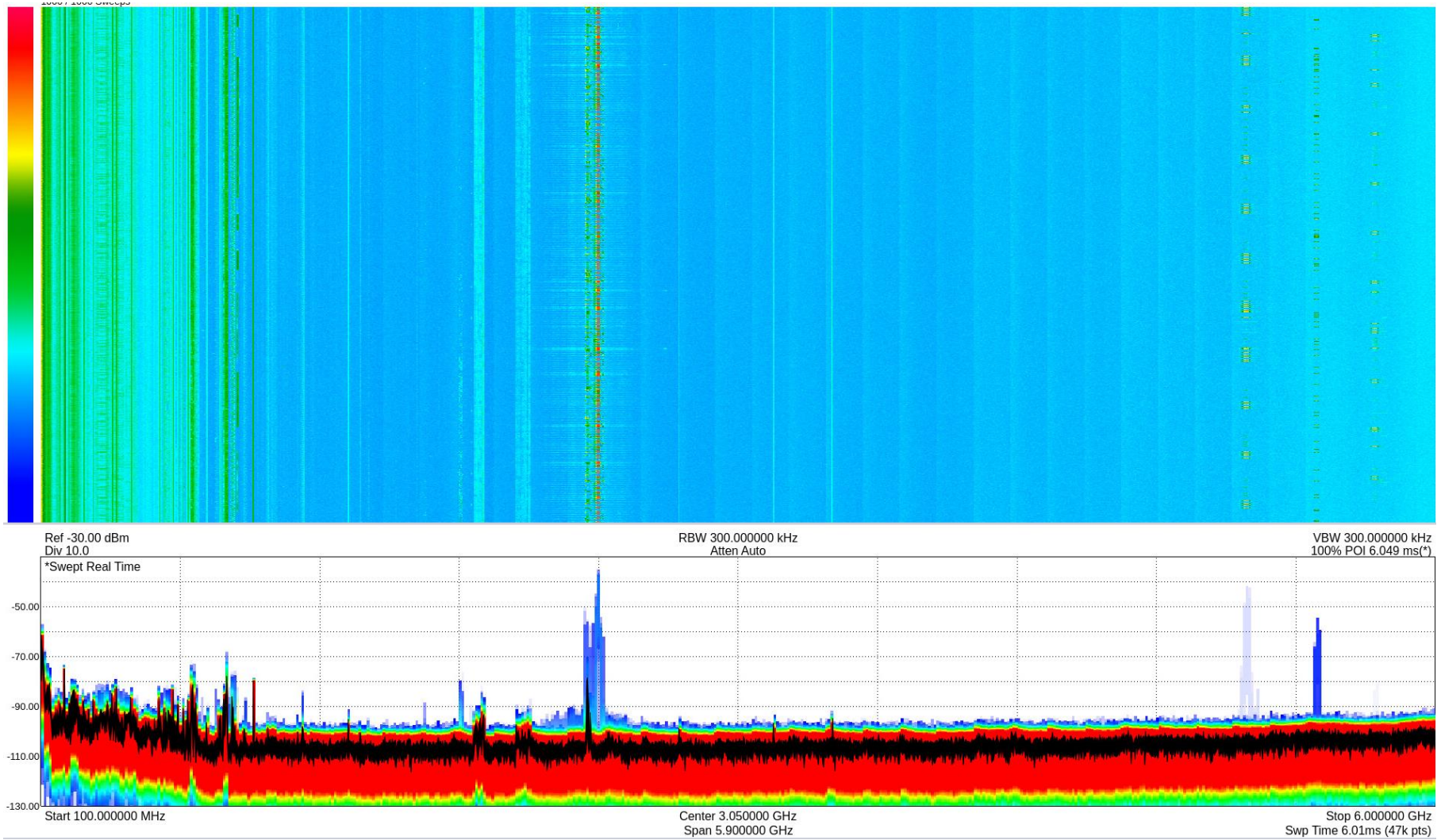
WCSNG
Wireless Communications Sensing and Networking

RFSynth: Data generation and testing platform for spectrum information systems

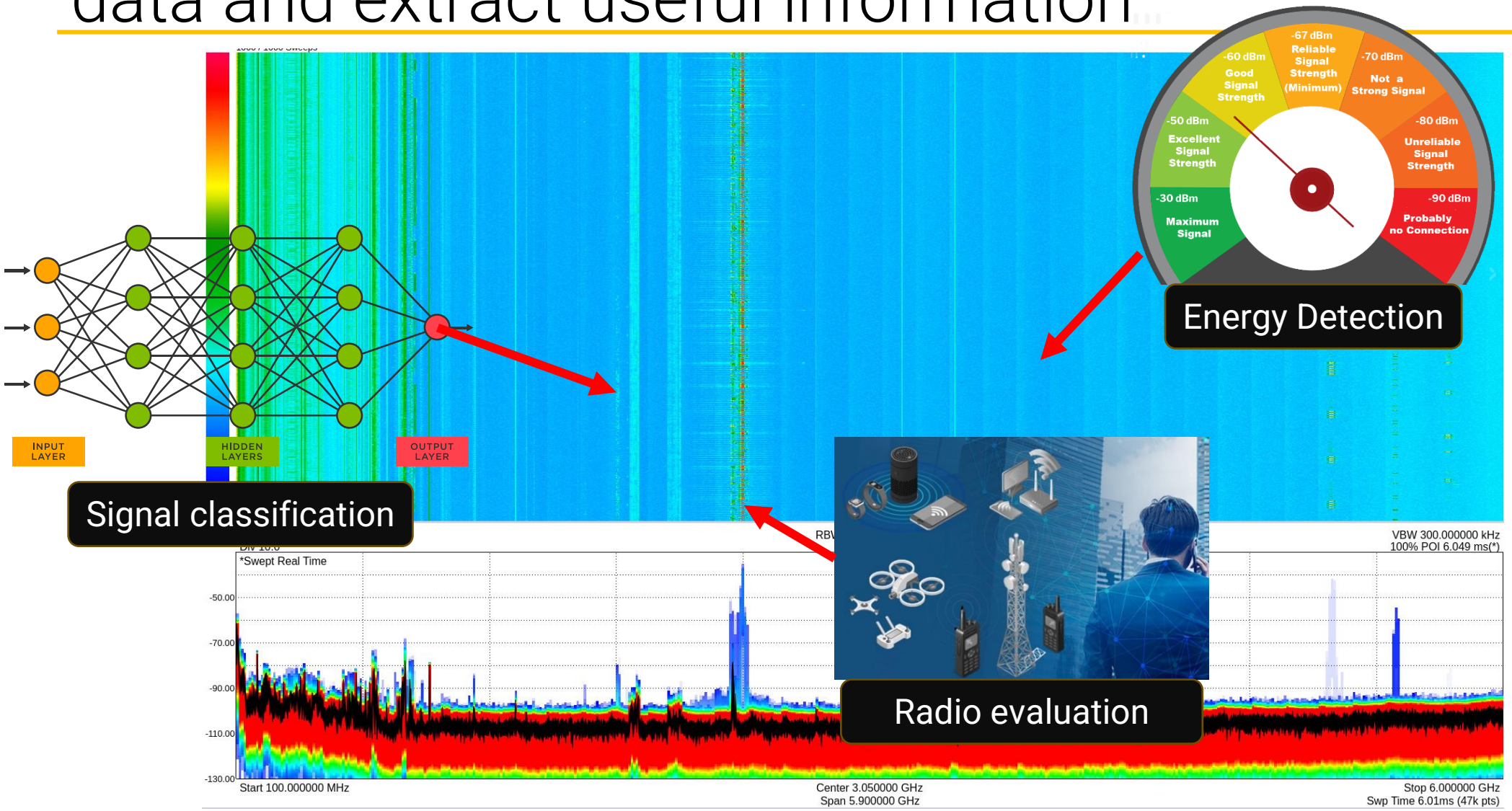
Raghav Subbaraman[*1], Hari Prasad Sankar[1*], Tianyi Hu[2], Dinesh Bharadia[1]

[1]University of California San Diego, [2]JASR Systems
*equally credited authors

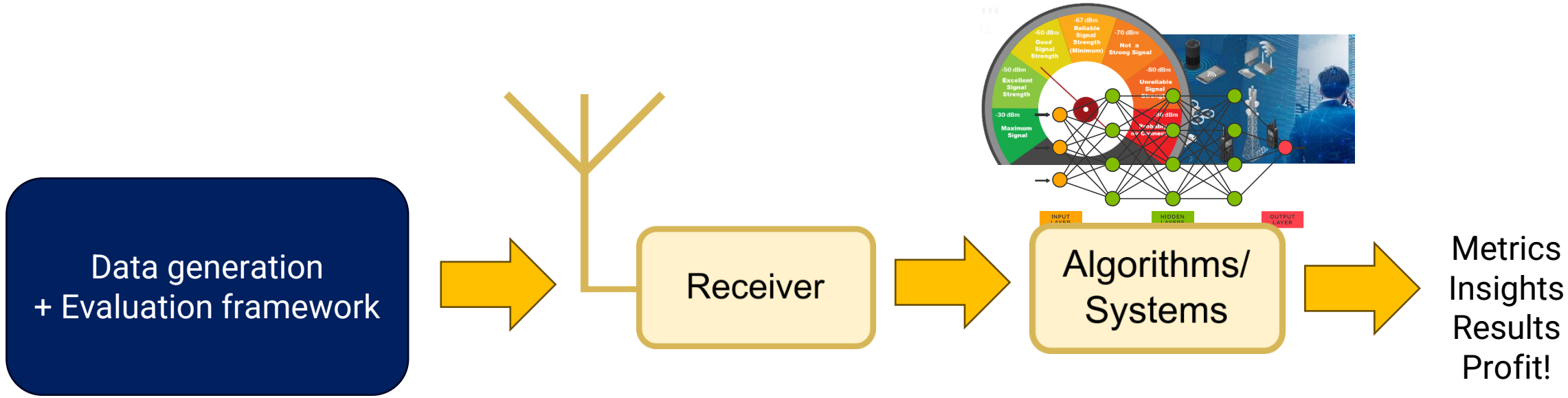
Spectrum information systems process spectrum data and extract useful information



Spectrum information systems process spectrum data and extract useful information



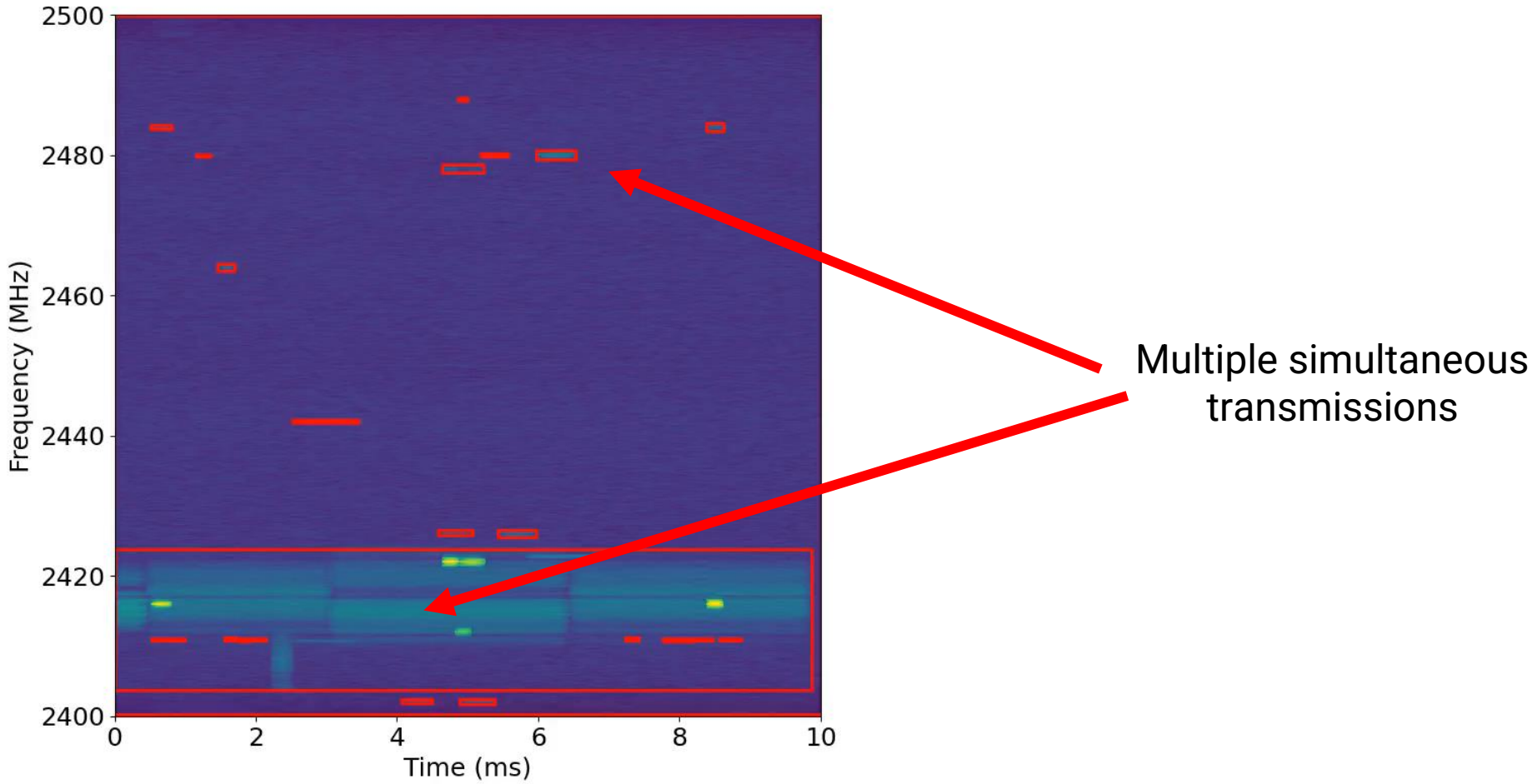
Data generation and testing for spectrum information applications is difficult!



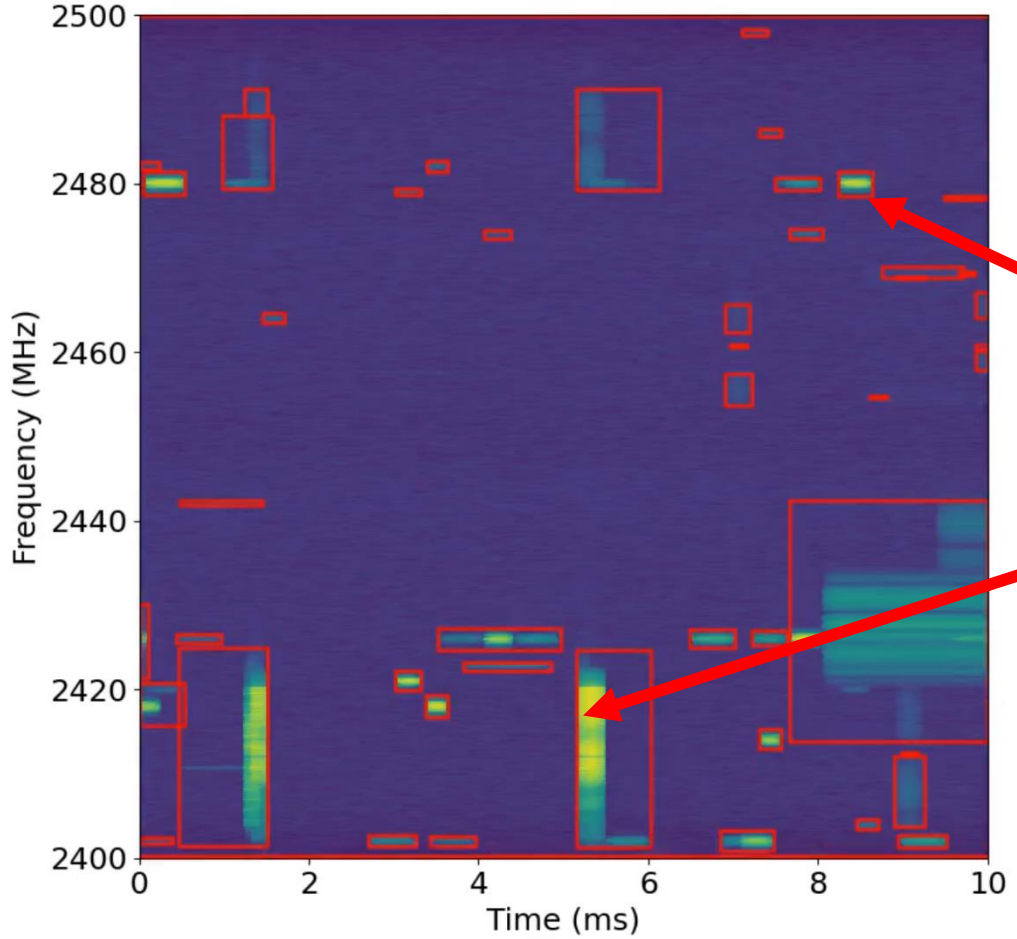
Only ~33% of NSF-SWIFT listed ML publications have available datasets!*

*thanks to Sangwon Shin, School of Computing UNL for compiling this metric

Wireless spectrum data should comprise of multiple dynamic transmitters

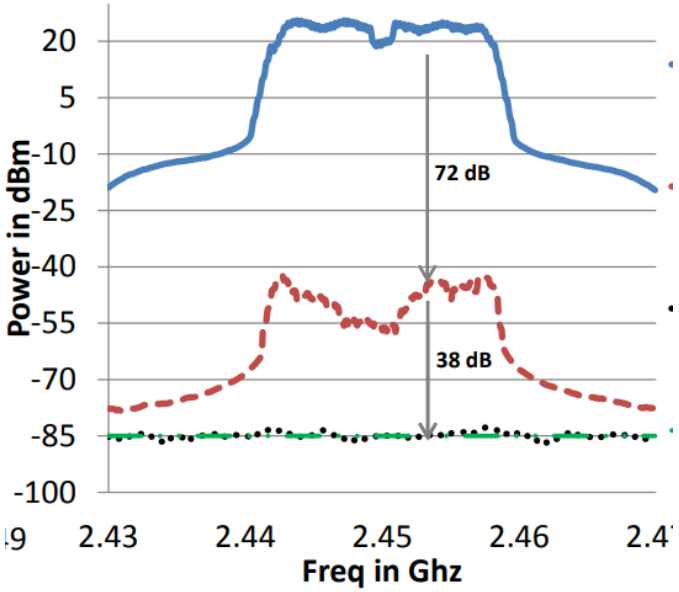


Context and metadata are a necessity



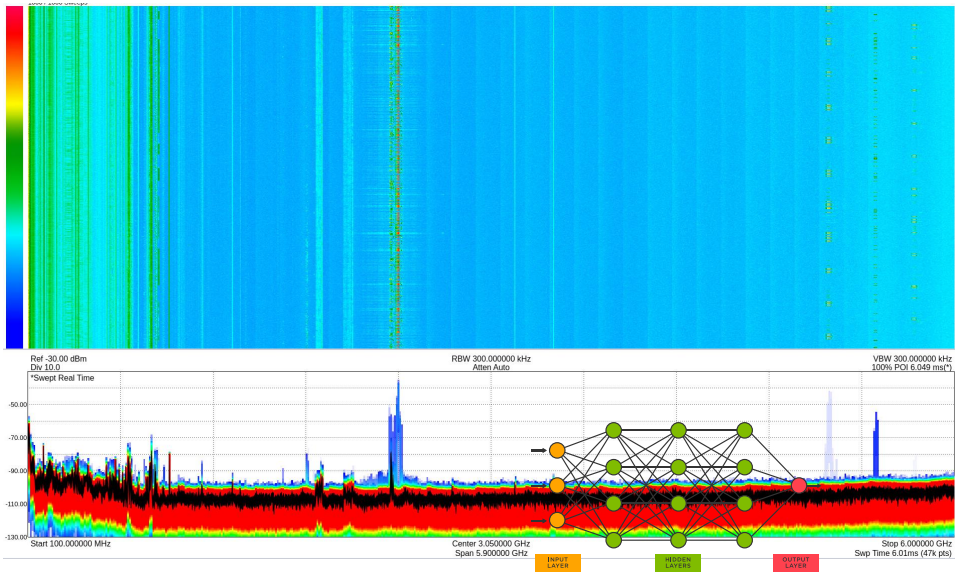
Who/what generated this?
How long was it on?
When was it on?
What type of signal is this?
…

Real-time, wireless evaluations help support radio dynamic applications



Wireless channel

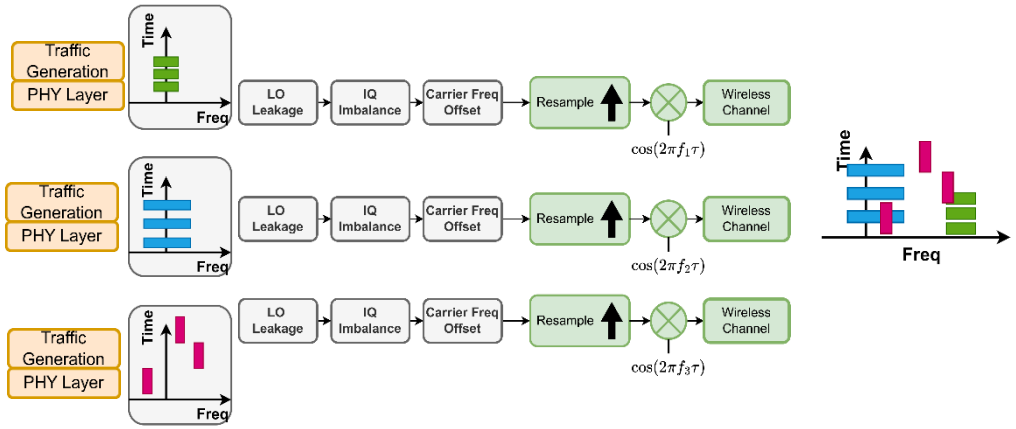
Live effects -- motion

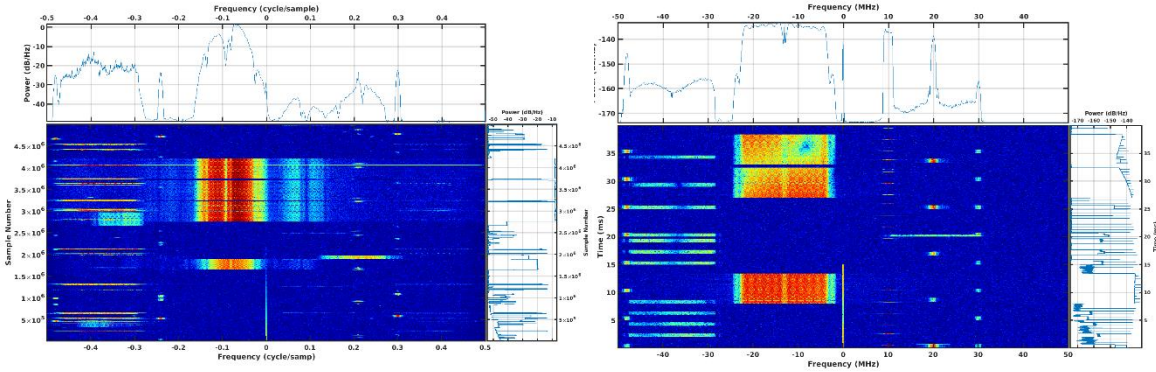


Test in real-RF environments
Respond to changes
Long-term studies

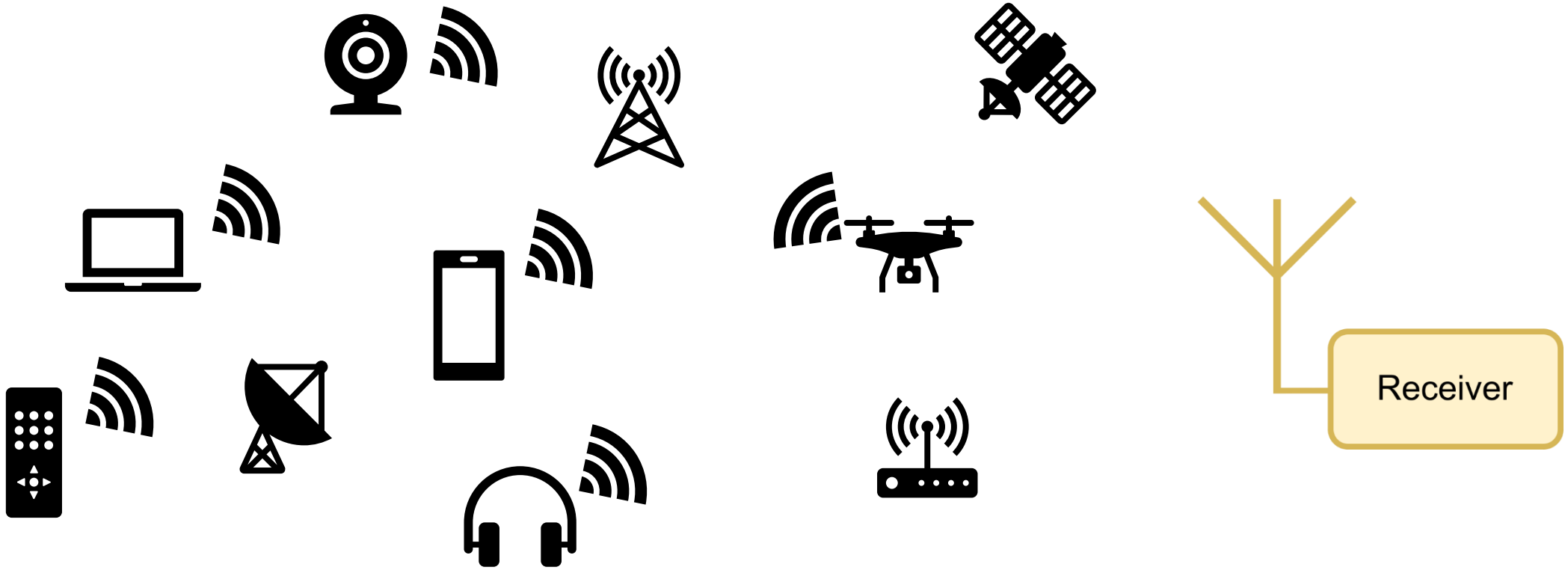
RFSynth: Our contributions

- System design for simultaneous multi-signal data generation
- Propose abstractions for metadata context
- Support for real-time, wireless evaluations
- Demo, data gallery, use-cases



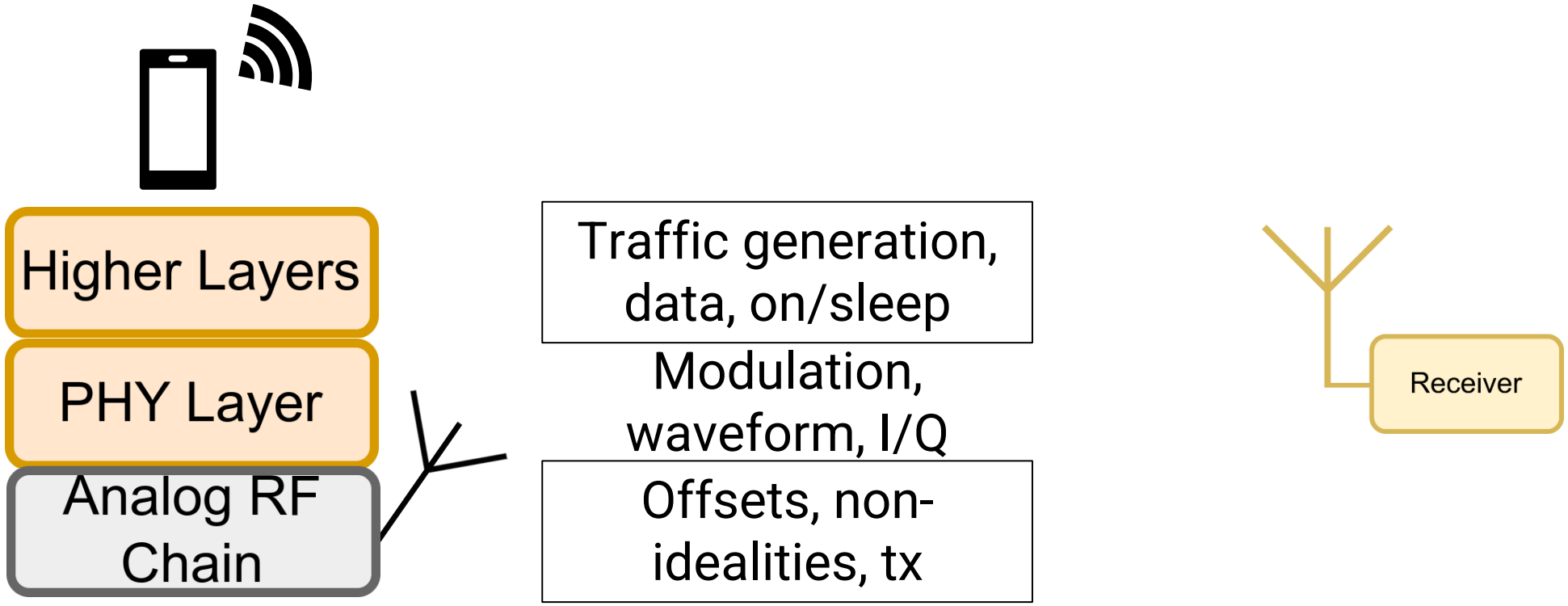


Generating representative RF data with many transmitters is hard!

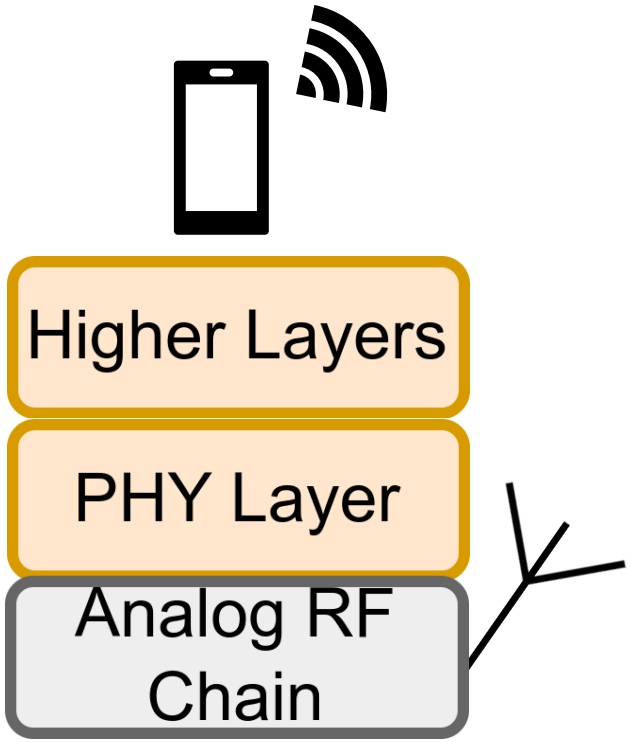


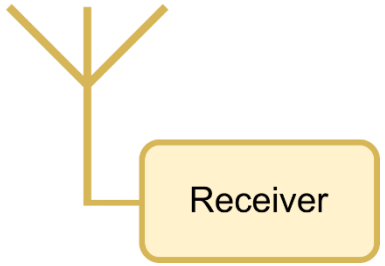
Controlling and orchestrating real devices is not always feasible

Layered approach to modeling simultaneous transmitters

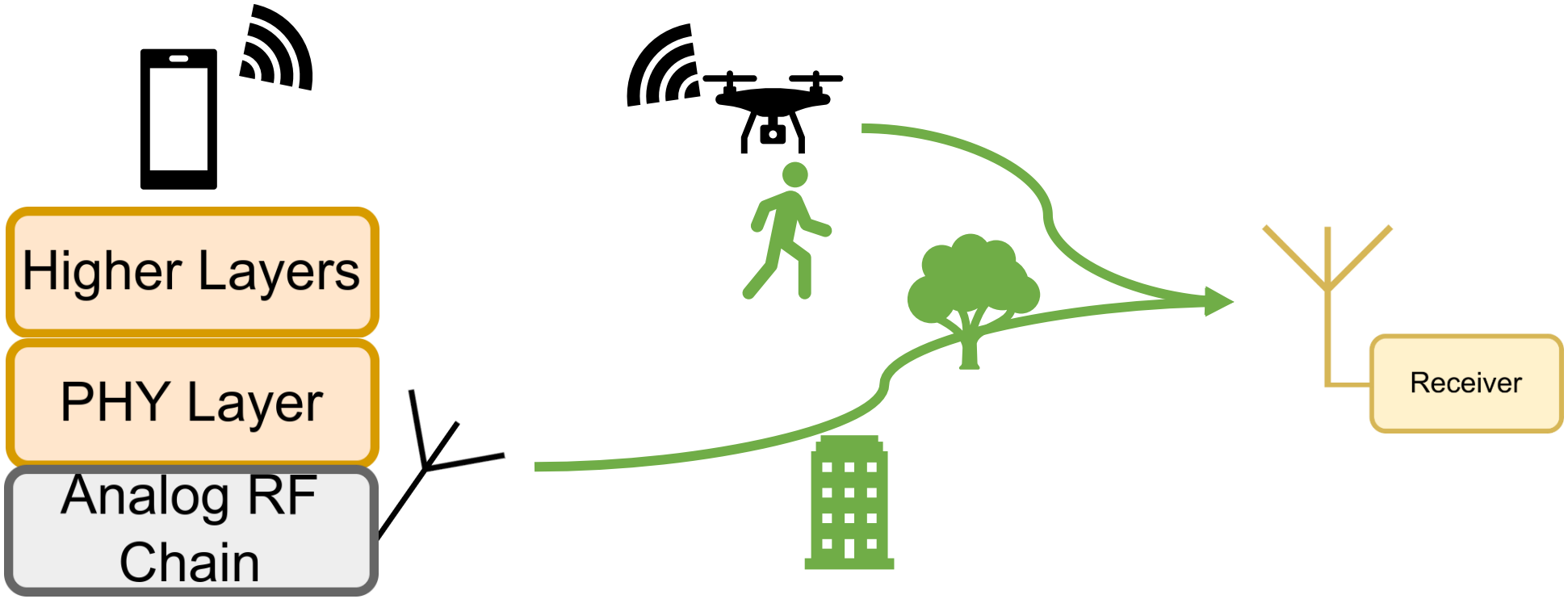


Not just the layers: the environment too!



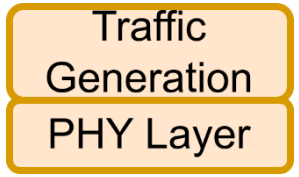


Not just the layers: the environment too!

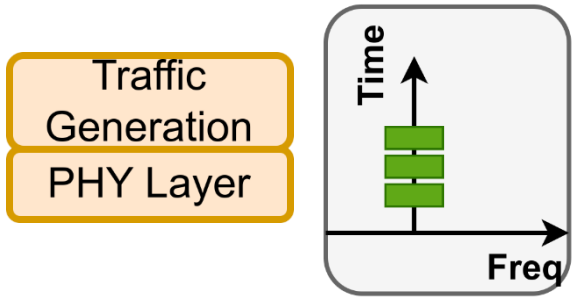


Environment:
Channel, superposition

Leverage layered approach to design data generation

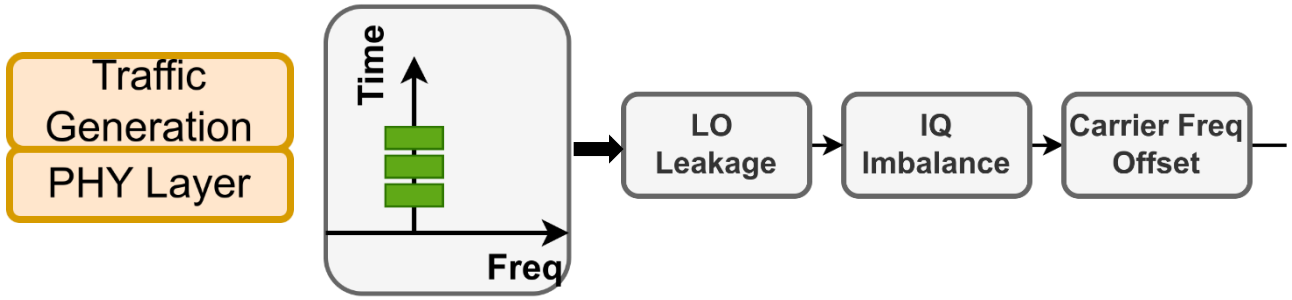


Leverage layered approach to design data generation



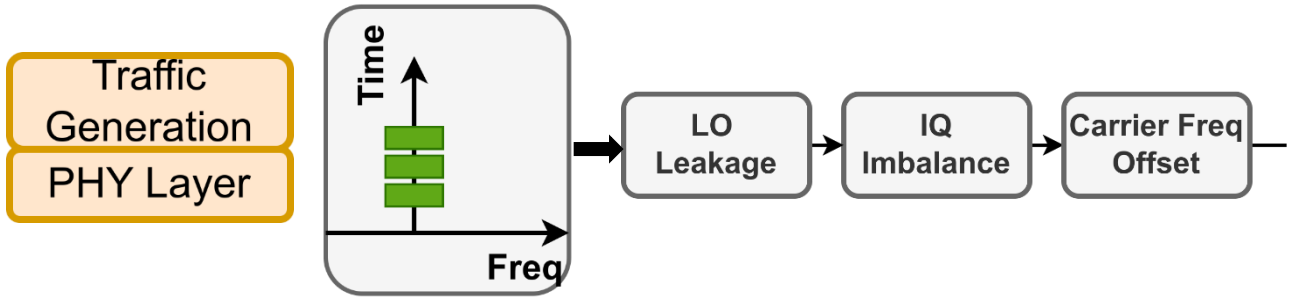
Baseband I/Q

Leverage layered approach to design data generation

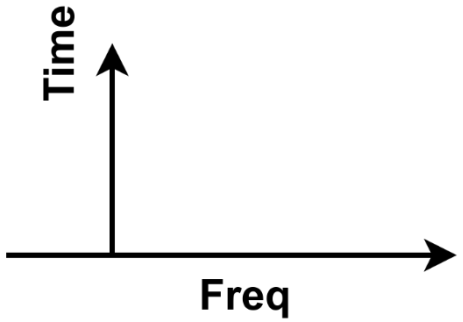


Baseband I/Q

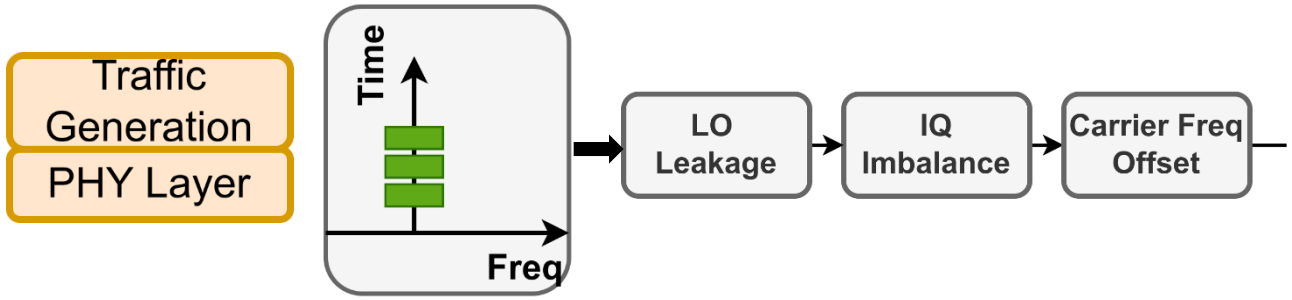
Leverage layered approach to design data generation



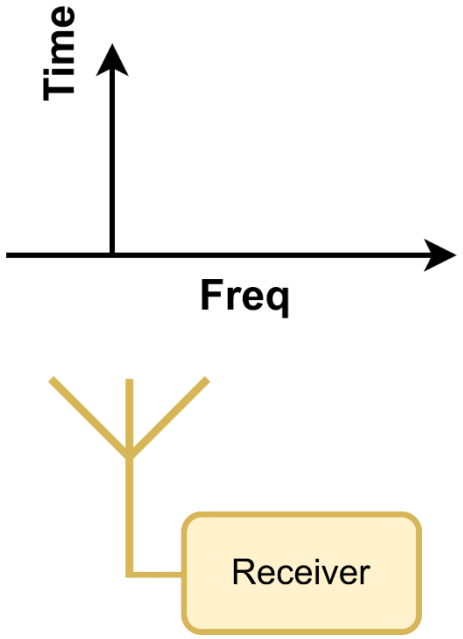
Baseband I/Q



Leverage layered approach to design data generation

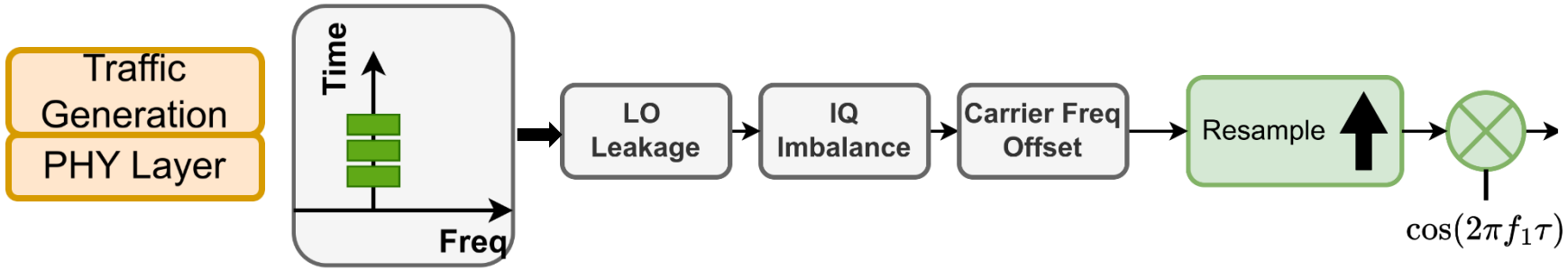


Baseband I/Q

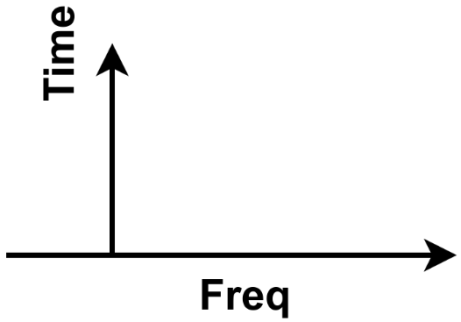


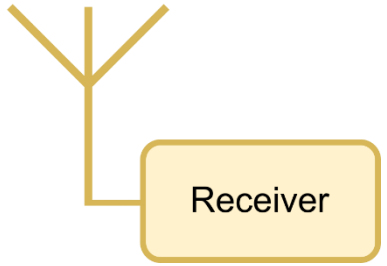
Baseband I/Q @ Receiver's sample rate

Leverage layered approach to design data generation



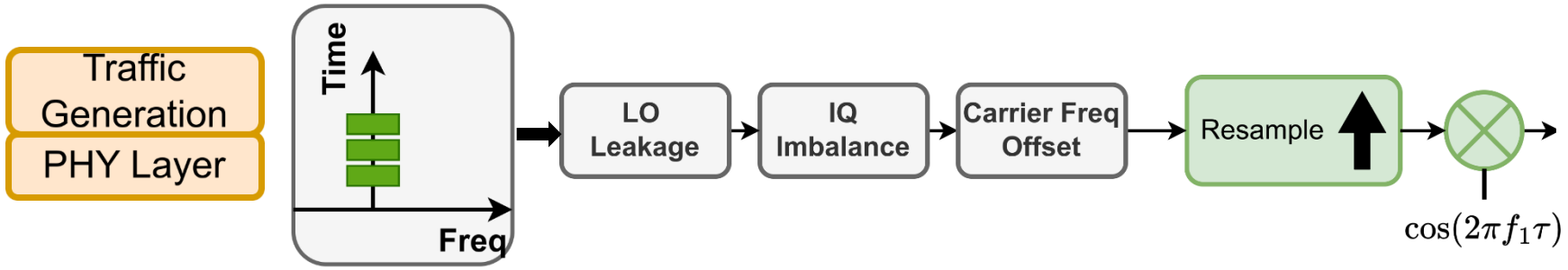
Baseband I/Q



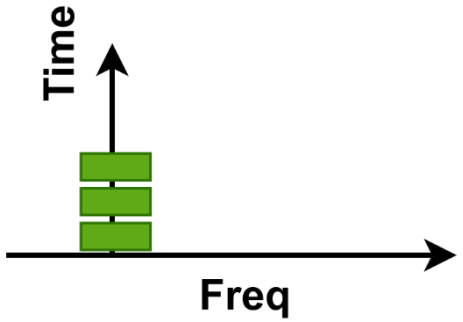


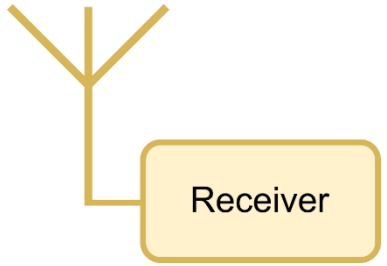
Baseband I/Q @ Receiver's sample rate

Leverage layered approach to design data generation



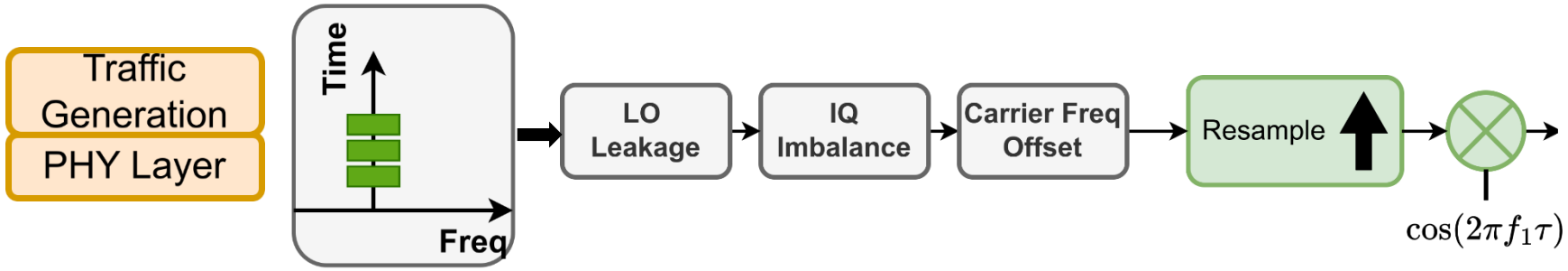
Baseband I/Q



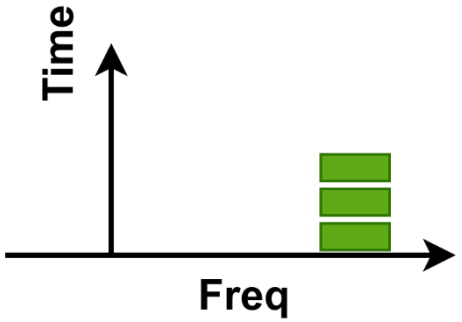


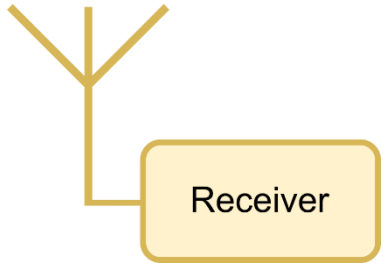
Baseband I/Q @ Receiver's sample rate

Leverage layered approach to design data generation



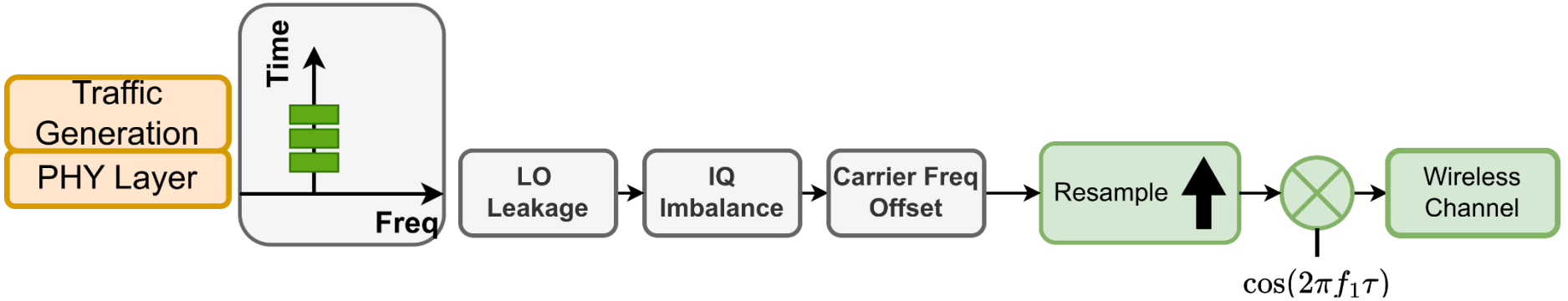
Baseband I/Q

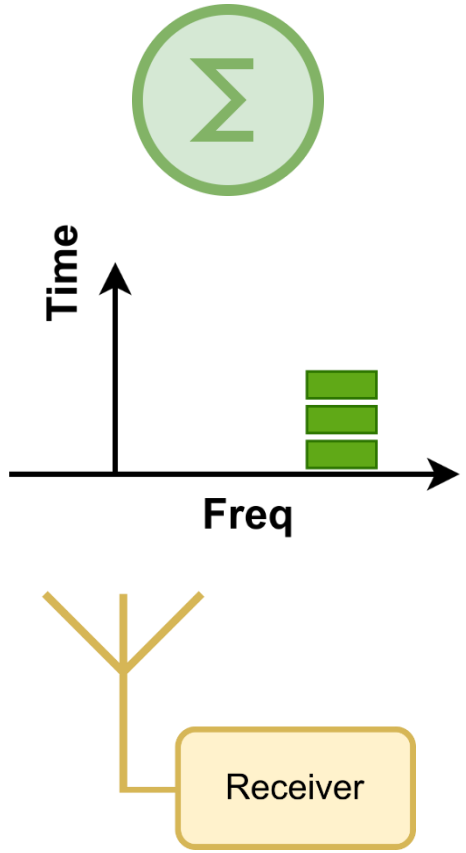




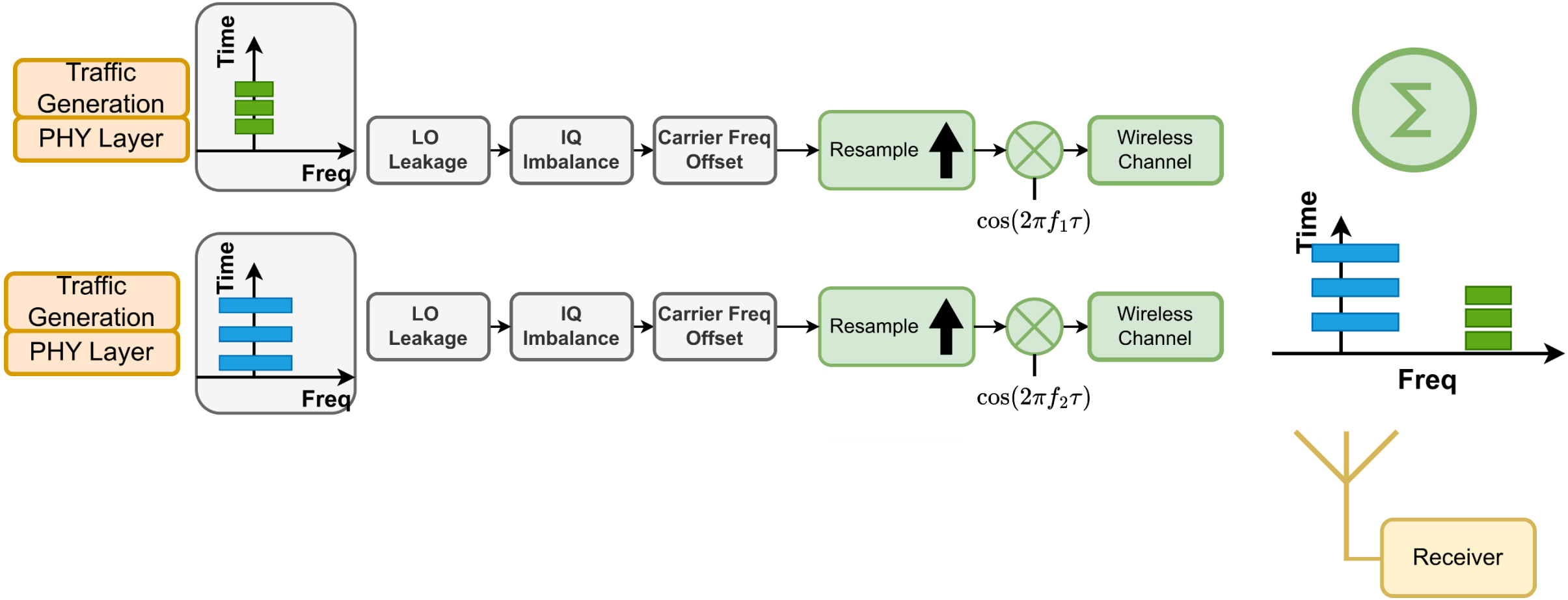
Baseband I/Q @ Receiver's sample rate

Extending to multiple simultaneous signals

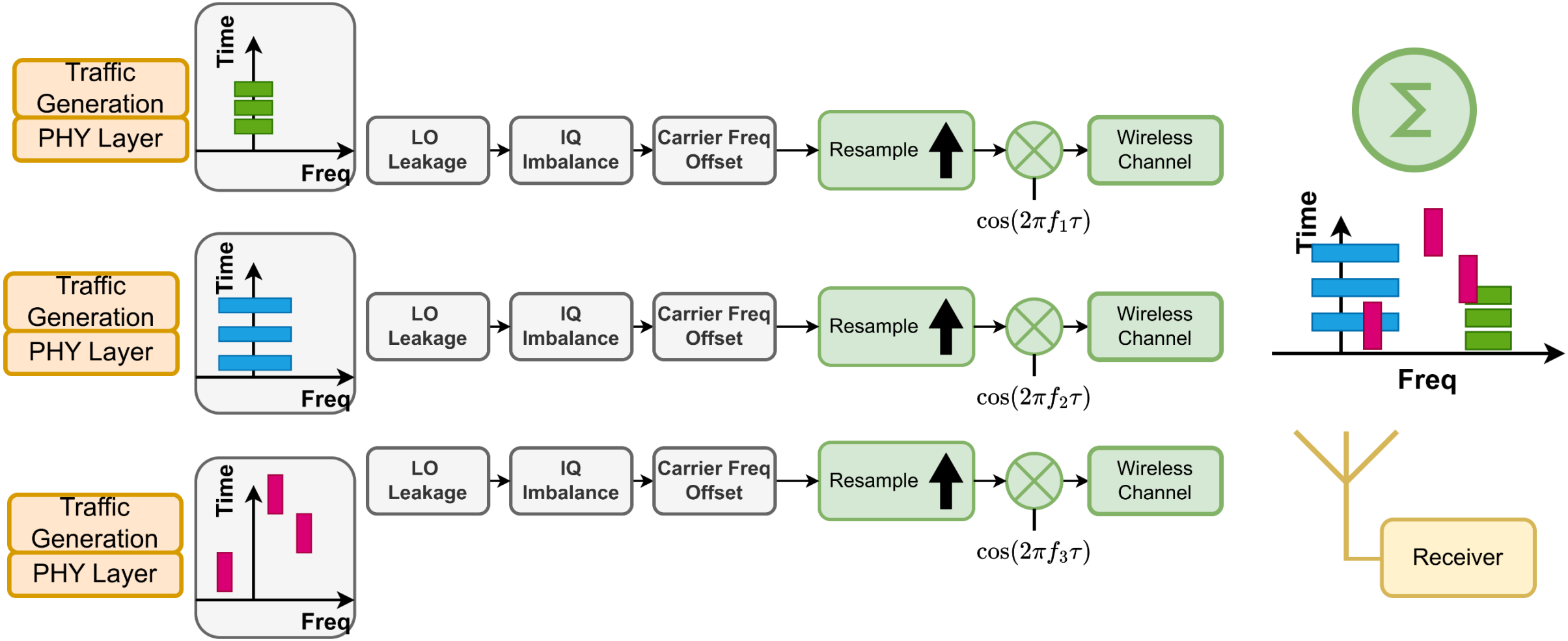




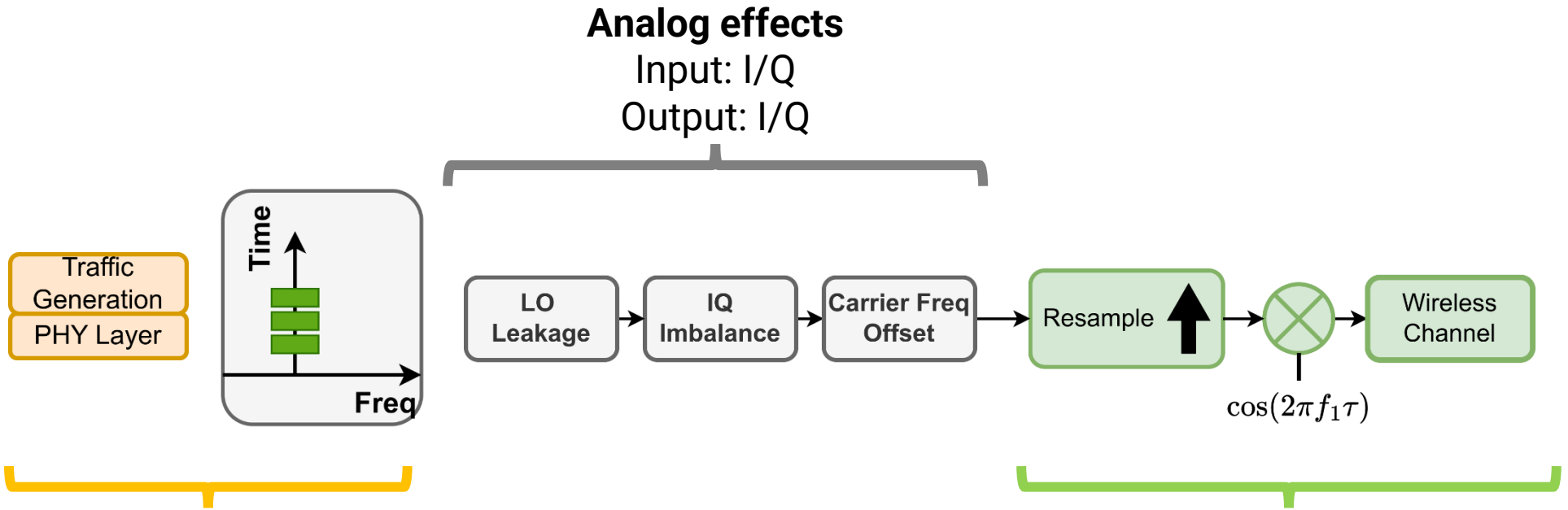
Extending to multiple simultaneous signals



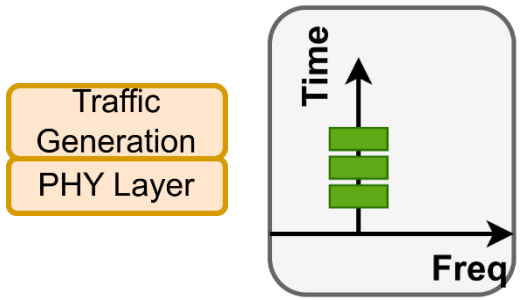
Extending to multiple simultaneous signals



Defining metadata abstractions



Defining metadata abstractions



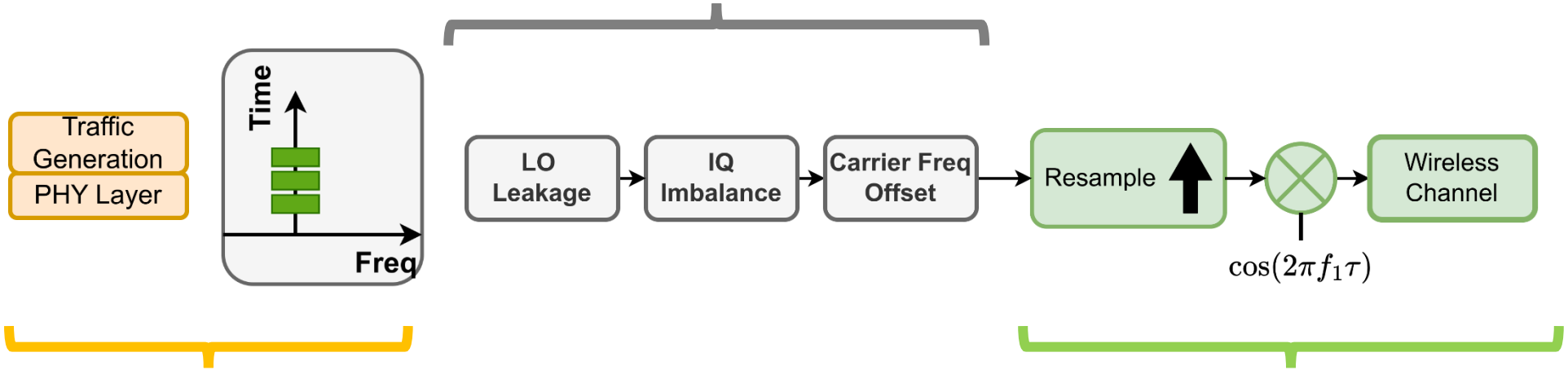
Signal Generator
Signal characteristics,
Transmission characteristics

“signal”
Conceptual unit

“energy”
One transmission

Defining metadata abstractions

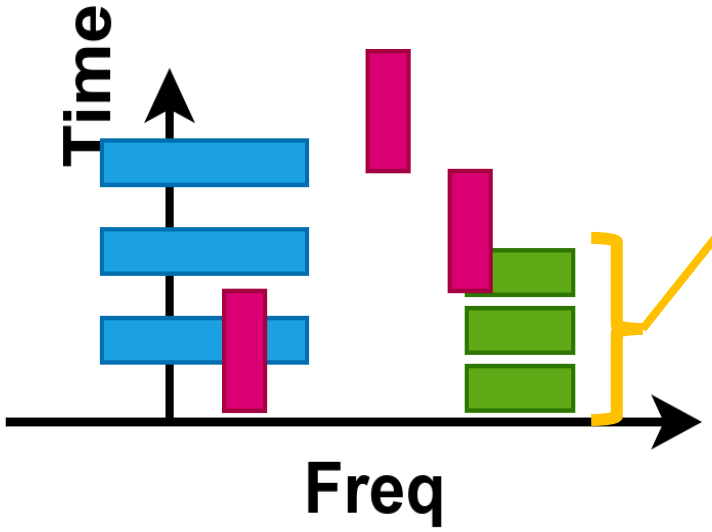
Analog effects – "source" metadata
Log transformations, update metadata



Signal Generator
Signal characteristics,
Transmission characteristics

Environment
Log frequency shifts, wireless channel

Example metadata



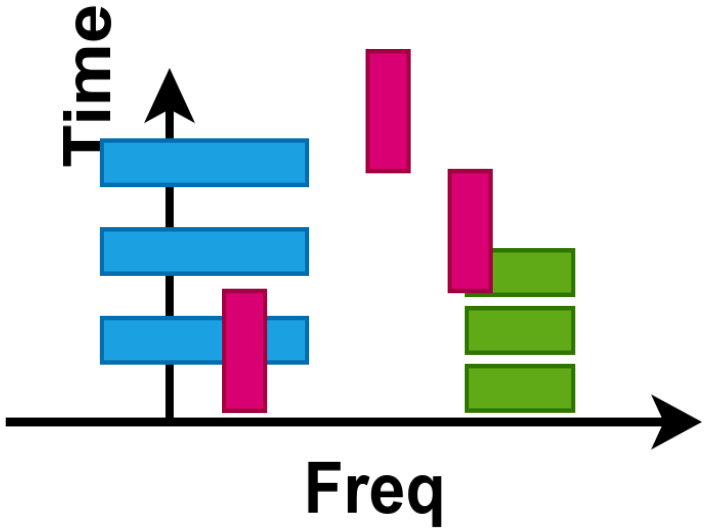

```
{
" report_type ": " signal " ,
" instance_name ": " bBiaaUMmSQmgIfLw0e " ,
" protocol ": " unknown " ,
" modality ": " single_carrier " ,
" modulation ": " qam16 " ,
" activity_type ": " overt baseline " ,
" time_start ": 1684952693.346253 ,
" time_stop ": 1684952803.348253 ,
" freq_lo ": 2481.875 ,
" freq_hi ": 2482.125 ,
" rx_center_freq ": {
" rx1 ": 2.482 E +9
} ,
" reference_time ": 65.001 ,
" reference_freq ": 2482 ,
" timeLength_s ": 110.002 ,
" bandwidth_Hz ": 0.25
```


Signal level metadata

Example metadata


```
{
" report_type ": " energy " ,
" instance_name ": " bBiaaUMmSQmgIfLw0e0EfE " ,
" time_start ": 1684952693.346253 ,
" time_stop ": 1684952693.348253 ,
" freq_lo ": 2481.875 ,
" freq_hi ": 2482.125 ,
" timeLength_s ": 0.0020000000000006679 ,
" bandwidth_Hz ": 0.25
}
```


Energy level metadata



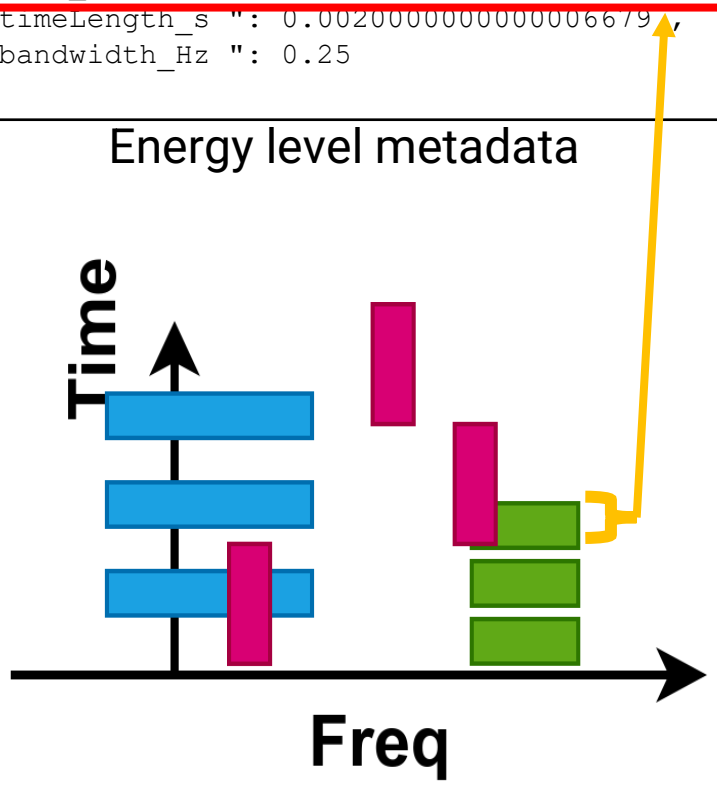

```
{
" report_type ": " signal " ,
" instance_name ": " bBiaaUMmSQmgIfLw0e " ,
" protocol ": " unknown " ,
" modality ": " single_carrier " ,
" modulation ": " qam16 " ,
" activity type ": " overt baseline " ,
" time_start ": 1684952693.346253 ,
" time_stop ": 1684952803.348253 ,
" freq_lo ": 2481.875 ,
" freq_hi ": 2482.125 ,
" rx_center_freq ": {
" rx1 ": 2.482 E +9
} ,
" reference_time ": 65.001 ,
" reference_freq ": 2482 ,
" timeLength_s ": 110.002 ,
" bandwidth_Hz ": 0.25
```


Signal level metadata

Example metadata


```
{
" report type ": " energy " ,
" instance_name ": " bBiaaUMmSQmgIfLw0e0EfE " ,
" time_start ": 1684952693.346253 ,
" time_stop ": 1684952693.348253 ,
" freq_lo ": 2481.875 ,
" freq_hi ": 2482.125 ,
" timeLength_s ": 0.0020000000000006679 ,
" bandwidth_Hz ": 0.25
}
```


Energy level metadata

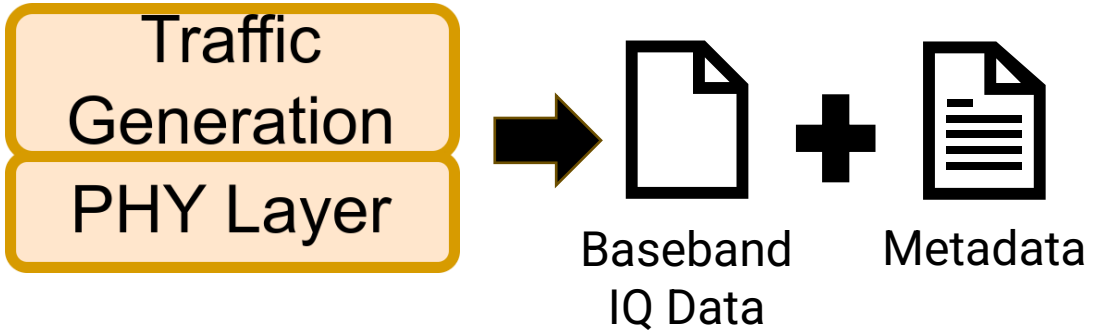



```
{
" report_type ": " signal " ,
" instance_name ": " bBiaaUMmSQmgIfLw0e " ,
" protocol ": " unknown " ,
" modality ": " single_carrier " ,
" modulation ": " qam16 " ,
" activity type ": " overt baseline " ,
" time_start ": 1684952693.346253 ,
" time_stop ": 1684952803.348253 ,
" freq_lo ": 2481.875 ,
" freq_hi ": 2482.125 ,
" rx_center_freq ": {
" rx1 ": 2.482 E +9
} ,
" reference_time ": 65.001 ,
" reference_freq ": 2482 ,
" timeLength_s ": 110.002 ,
" bandwidth_Hz ": 0.25
```

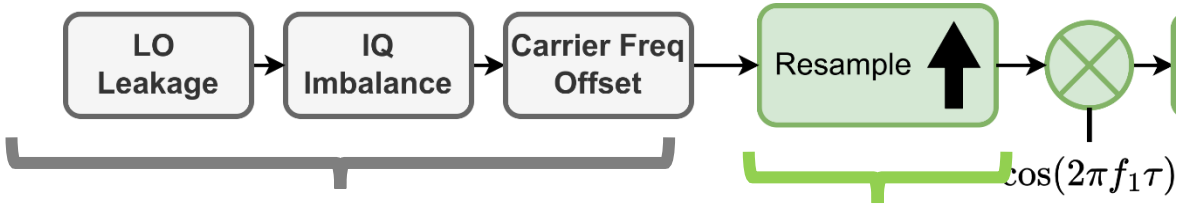

Signal level metadata

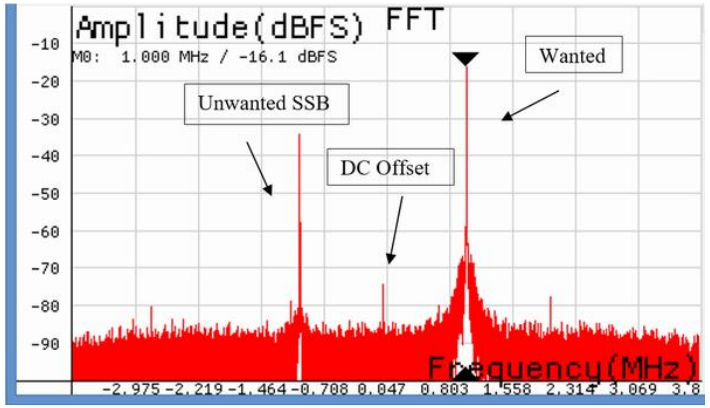
Signal generation support

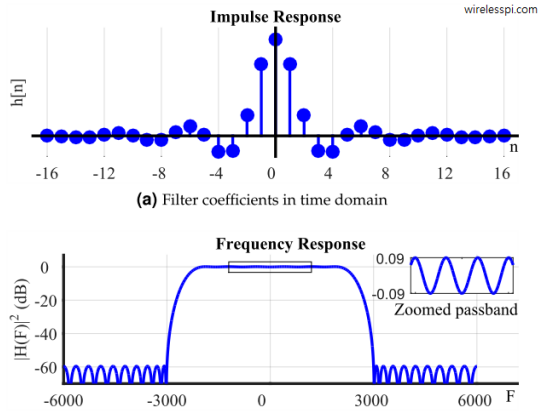
Modulation only	Protocol Compliant	Anomalous Signals
FSK family – FSK, GFSK, MSK,GMSK	LTE	Spread spectrum interference
PSK family	5G NR	Sinusoidal spurs
QAM family	Zigbee	RF Emanations
Analog Modulations – AM/FM	Wi-Fi	Gaussian noise bursts
OFDM	LoRa	
PAM, OOK	BLE	



Analog artifacts and environment – implementation considerations







$$\begin{cases} \tilde{x}_I(t) &= (1+\varepsilon_A)[x_I(t)\cos(\varepsilon_\theta/2) - x_Q(t)\sin(\varepsilon_\theta/2)] \\ \tilde{x}_Q(t) &= (1-\varepsilon_A)[x_Q(t)\cos(\varepsilon_\theta/2) - x_I(t)\sin(\varepsilon_\theta/2)] \end{cases}$$

Channel models

Channel Model	Model representation*
Identity	h(n)=1, ∀ n
Rician	$h_\ell[m] = \sqrt{\frac{\kappa}{\kappa+1}}\sigma_\ell e^{j\theta} + \sqrt{\frac{1}{\kappa+1}}\mathcal{CN}(0,\sigma_\ell^2)$
Free space pathloss	L = 20*log10(4πR/λ)
Rayleigh	$\lvert h[m]\rvert^2$ follows the distribution, $f(x) = \frac{1}{\sigma_\ell^2}\exp\left\{\frac{-x}{\sigma_\ell^2}\right\},\ x \geq 0$

*Fundamentals of wireless communication by David Tse & Pramod Vishwanath

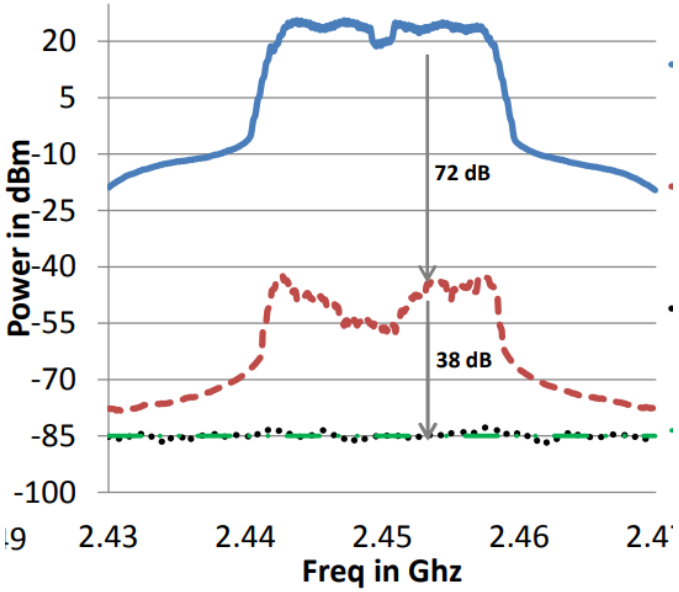
Channel models

Channel Model	Model representation*
Identity	h(n)=1, ∀ n
Rician	$h_\ell[m] = \sqrt{\frac{\kappa}{\kappa+1}}\sigma_\ell e^{j\theta} + \sqrt{\frac{1}{\kappa+1}}\mathcal{CN}(0,\sigma_\ell^2)$
Free space pathloss	L = 20*log10(4πR/λ)
Rayleigh	$\lvert h[m]\rvert^2$ follows the distribution, $f(x) = \frac{1}{\sigma_\ell^2}\exp\left\{\frac{-x}{\sigma_\ell^2}\right\},\ x \geq 0$

New channel models are plug-and play within the modular design

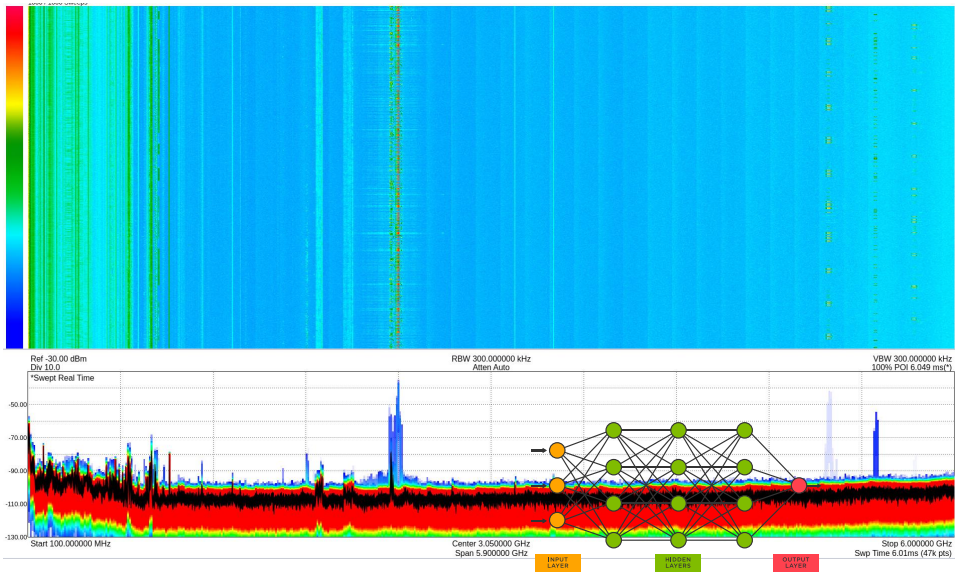
*Fundamentals of wireless communication by David Tse & Pramod Vishwanath

Enabling, wireless, real-time evaluation



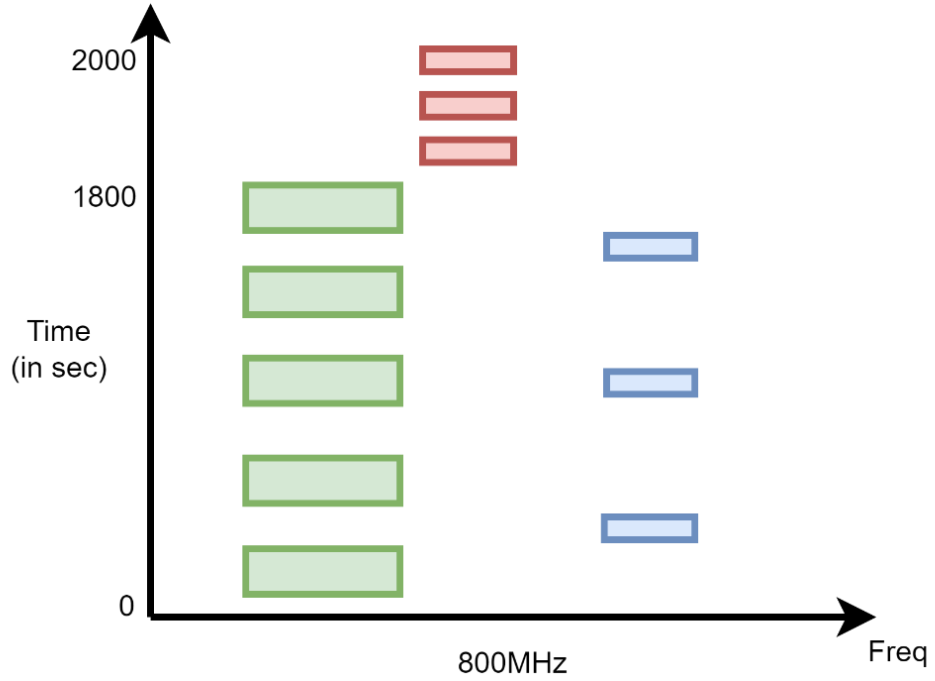
Wireless channel

Live effects -- motion



Test in real-RF environments
Respond to changes
Long-term studies

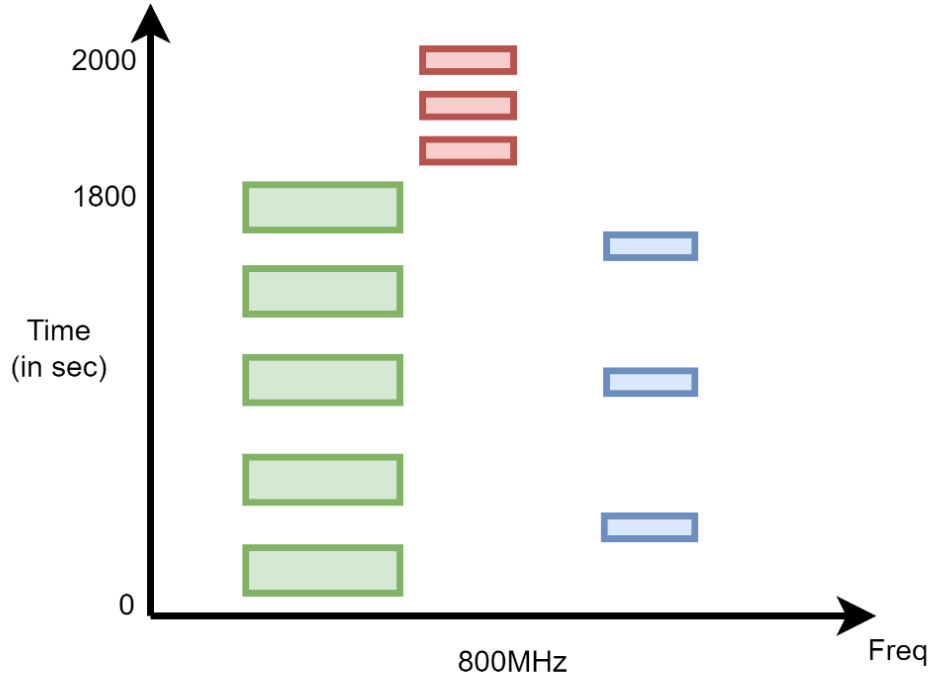
I/Q generation does not scale to real-time, long time-scale testing



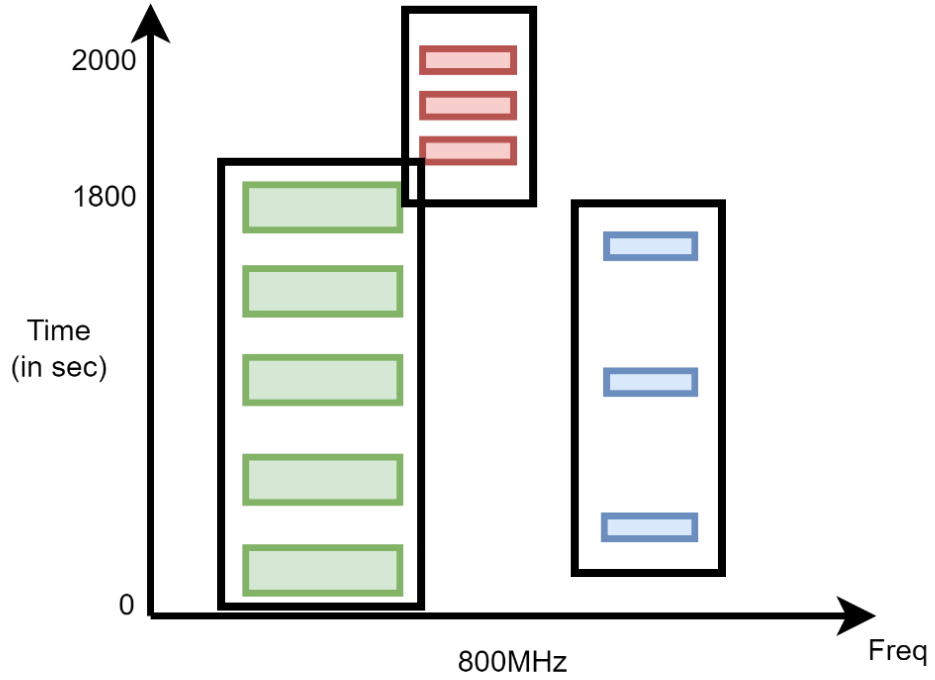
Total Time of IQ to be generated	2000 secs
Sample Rate	100e6
Total Sample generated	2e11 samples
Total file size	1600 GB

I/Q file sizes scale linearly with time and observation bandwidth

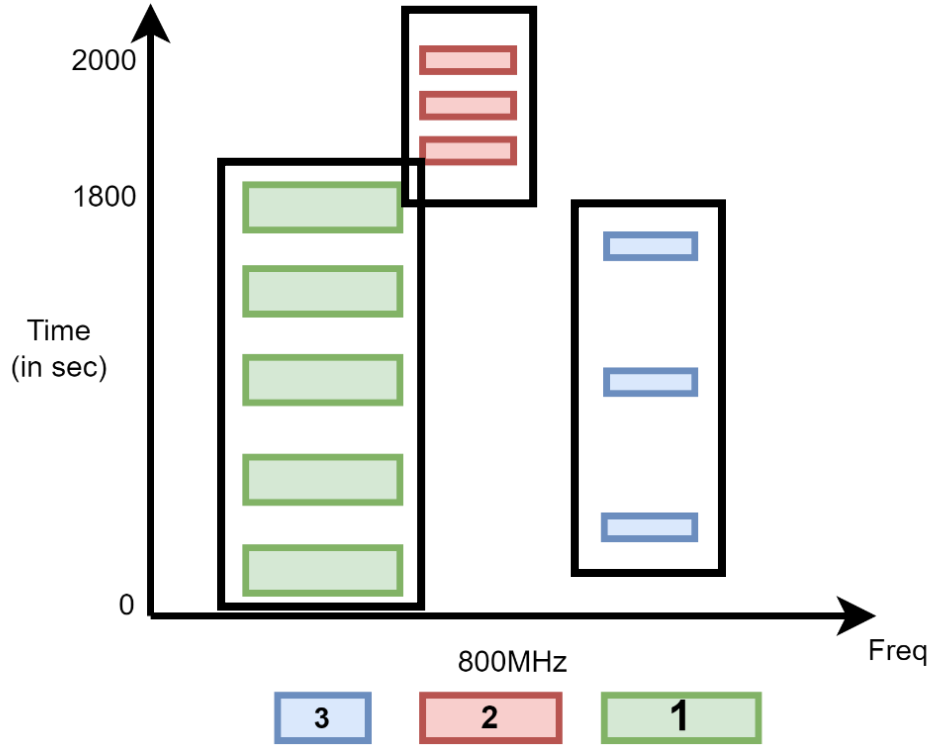
Insight: signals can be “compressed”



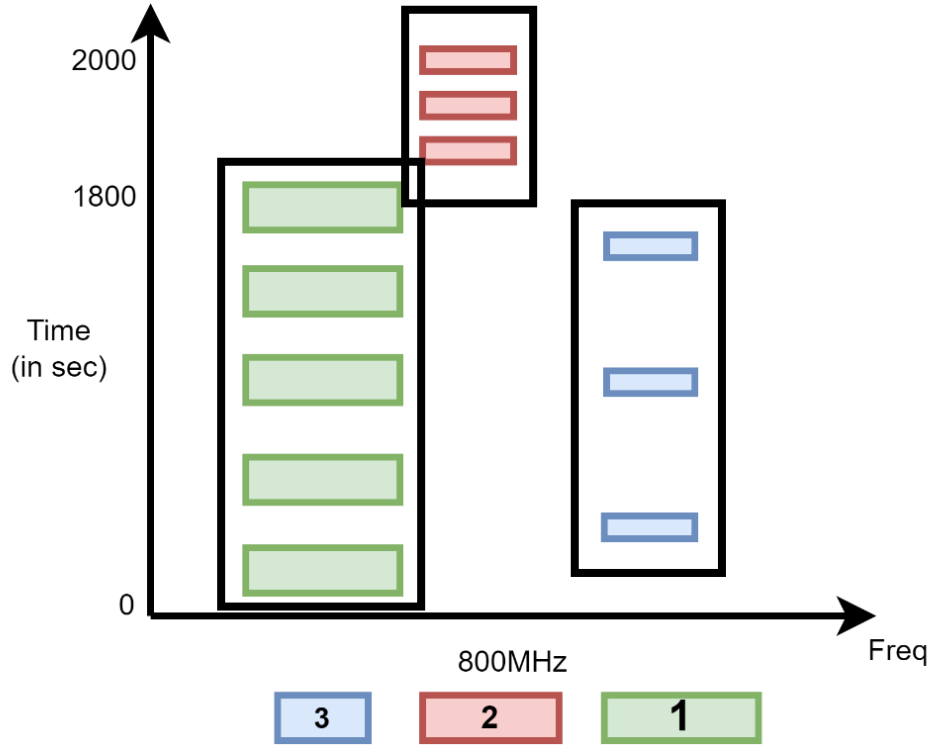
Insight: signals can be “compressed”



Insight: signals can be “compressed”



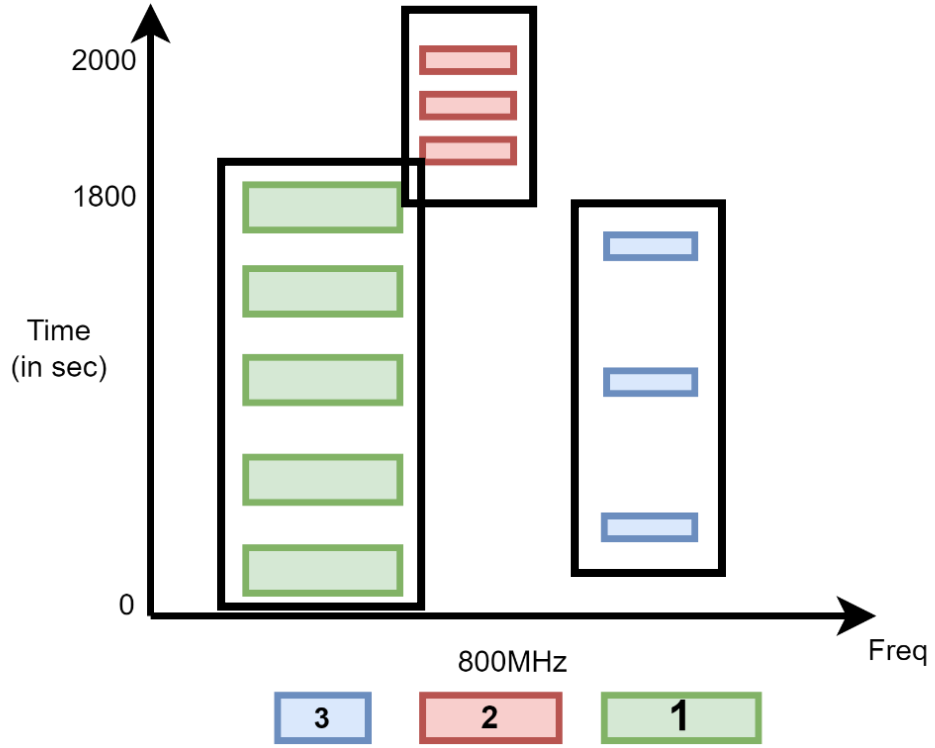
Insight: signals can be “compressed”



Total Time of IQ to be generated	2000 secs
Sample Rate	100e6
Total complex IQ Samples generated	2e11 samples
Total file size (compressed)	**80 MB!**

Assumptions:
Packet duration is 100ms
Total IQ samples to be generated: 10e6

Insight: signals can be “compressed”

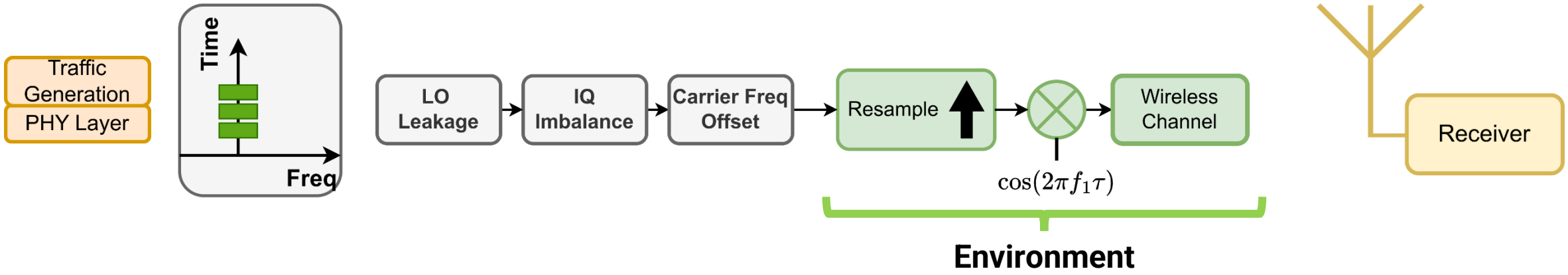


Total Time of IQ to be generated	2000 secs
Sample Rate	100e6
Total complex IQ Samples generated	2e11 samples
Total file size (compressed)	**80 MB!**

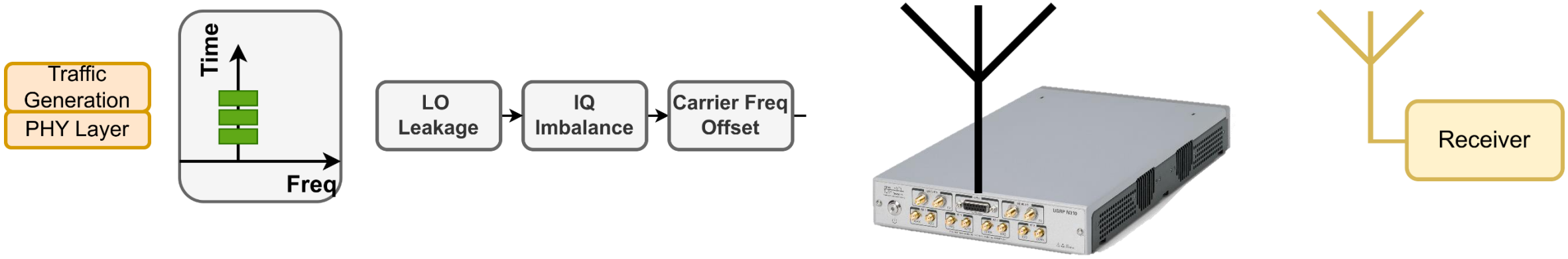
Assumptions:
Packet duration is 100ms
Total IQ samples to be generated: 10e6

Approximate signals as repeating bursts at precise time intervals! -> constant I/Q size

Connecting RFSynth to the wireless world through SDRs

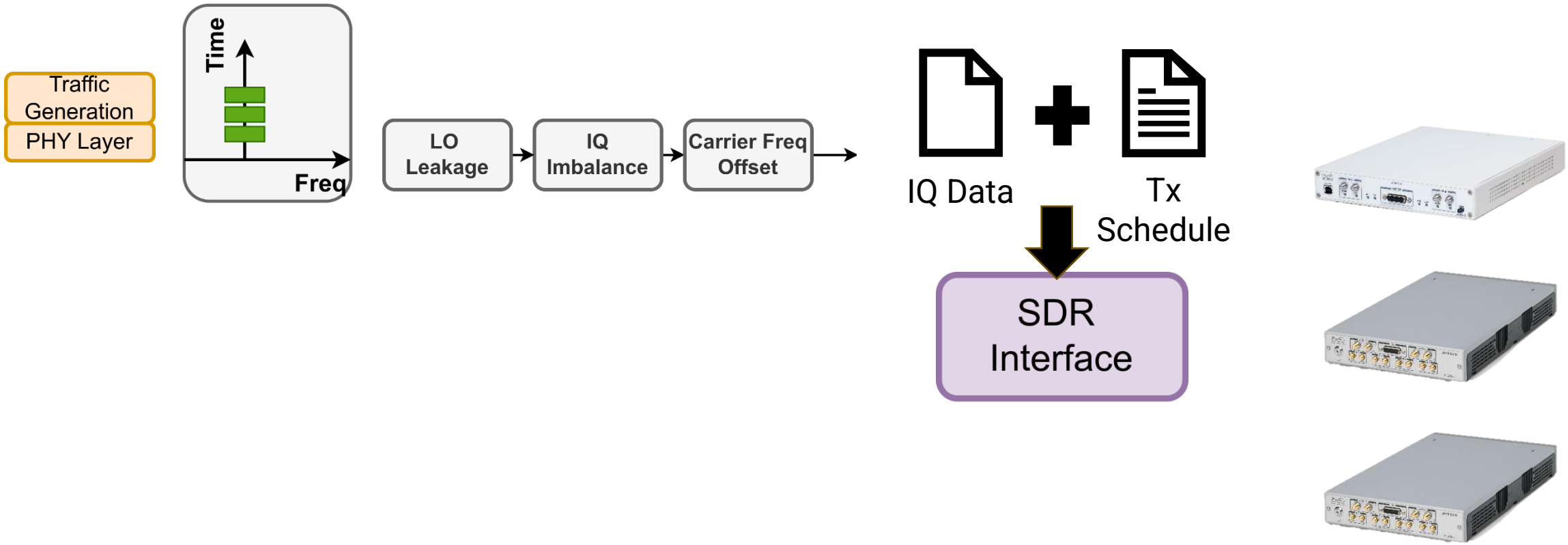


Connecting RFSynth to the wireless world through SDRs

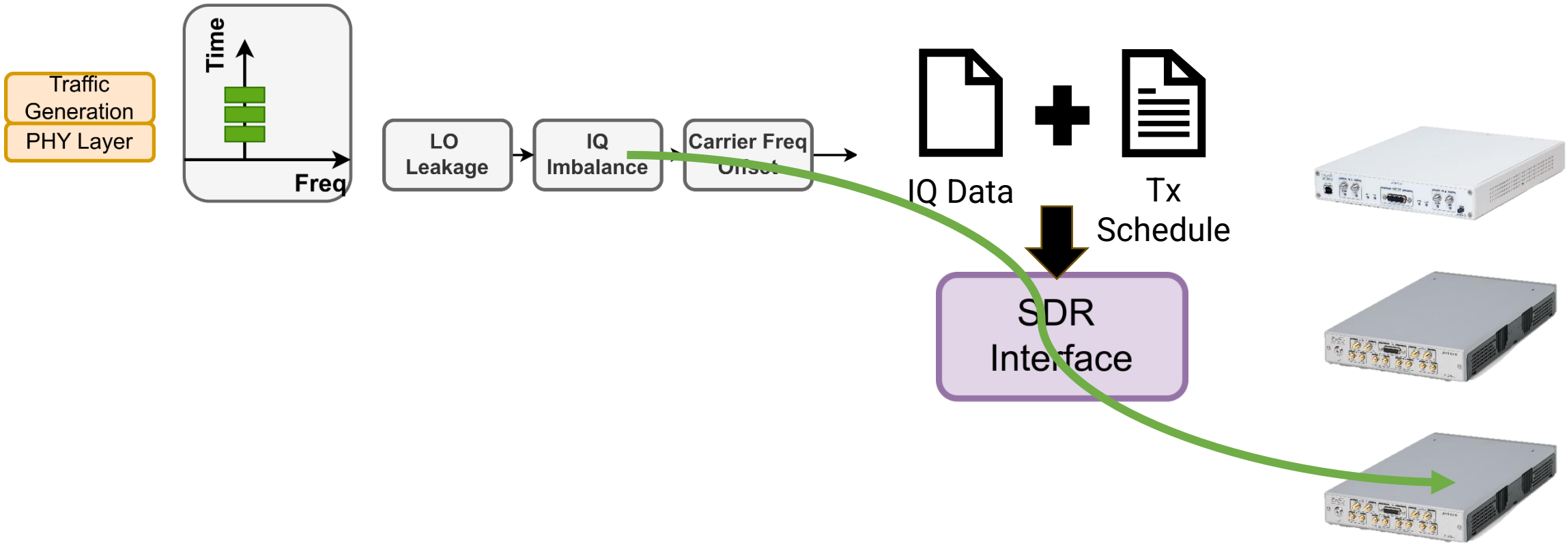


SDRs allow the transmitters to directly interface with the real world environment

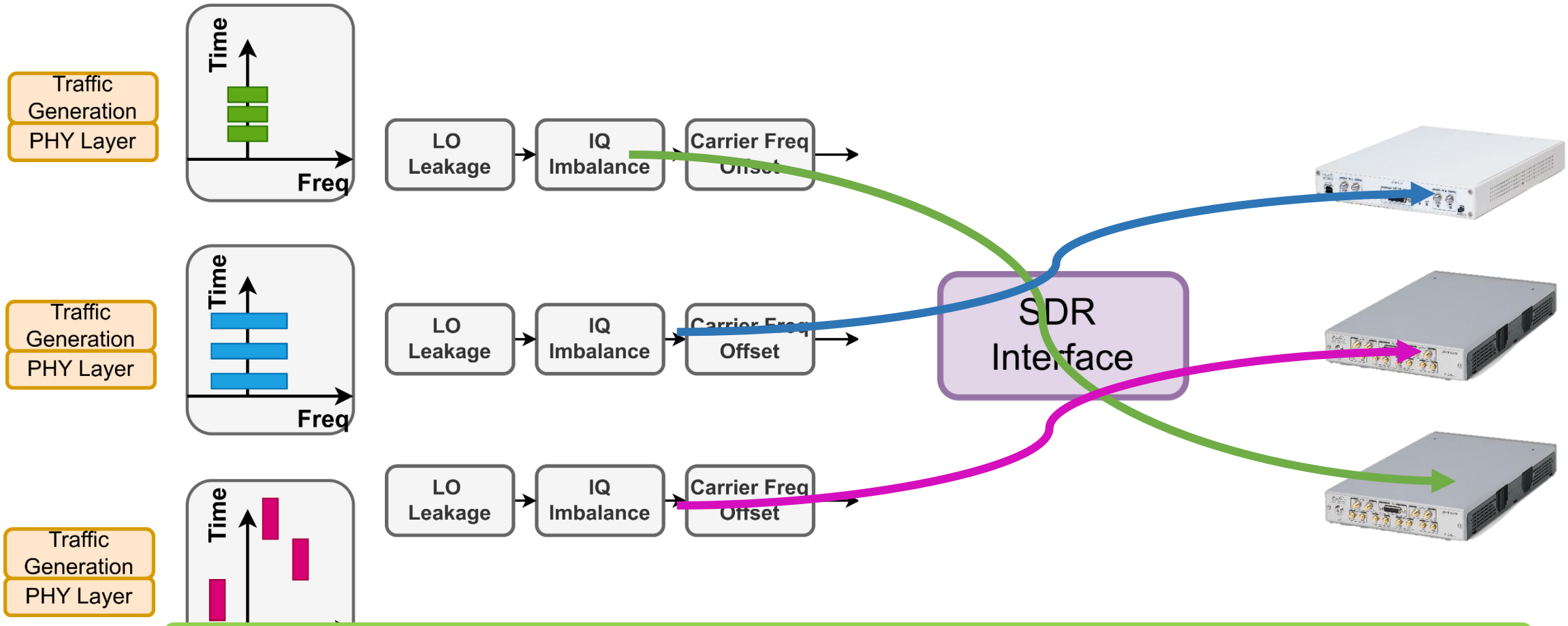
SDR interface dynamically maps signals to transmitters



SDR interface dynamically maps signals to transmitters

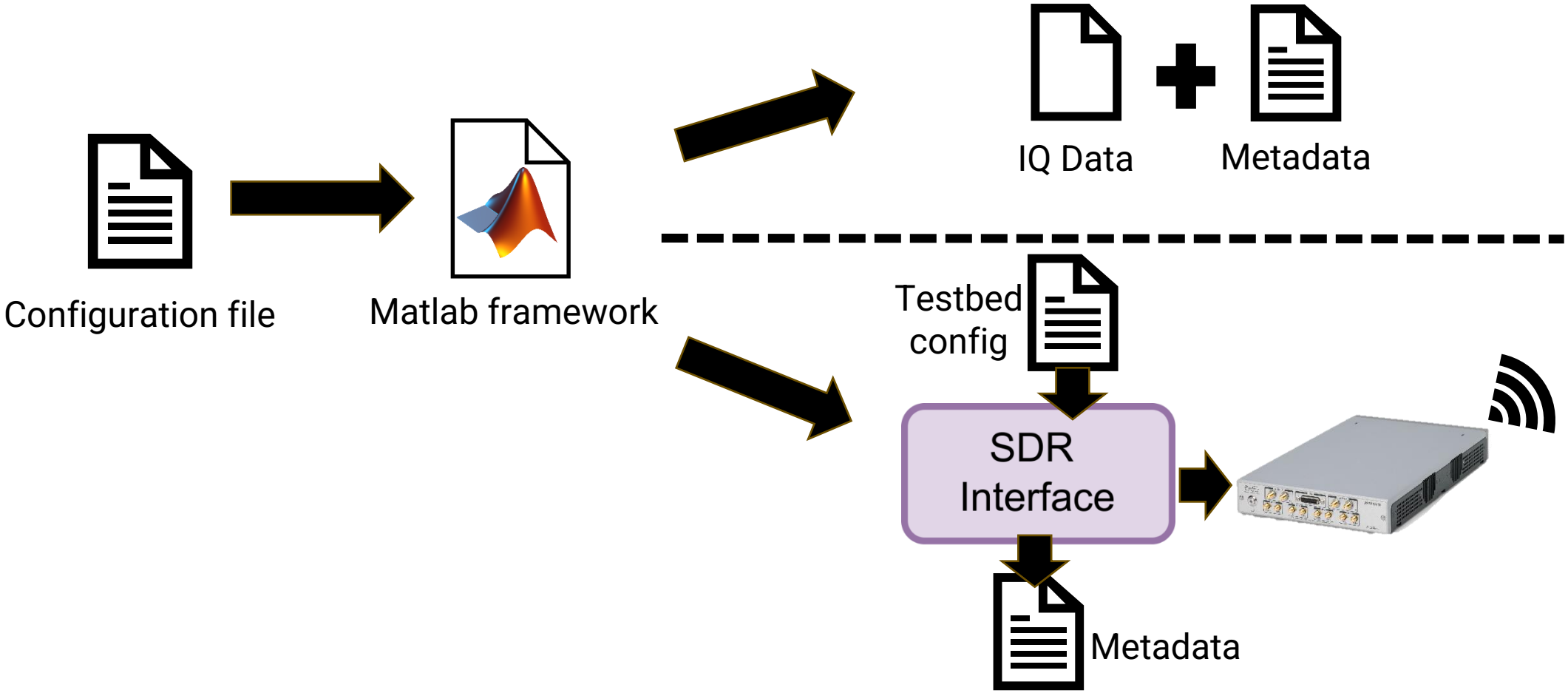


SDR interface dynamically maps signals to transmitters



SDRs are resources used on a need basis – signals can turn on-off

RFSynth usage flow

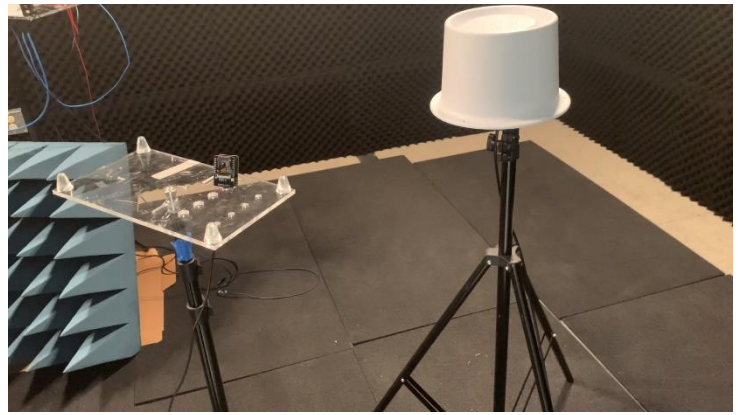


Demo video

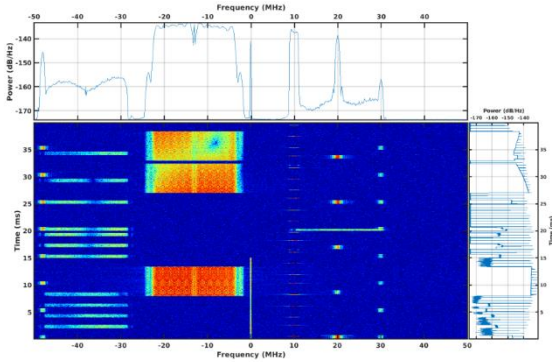
YouTube Link: https://www.youtube.com/watch?v=inNMCq5oWZ0

RFSynth @ UCSD Signals Lab

Lab-space with 5x N320 + 3x SM200C + 4x X300, GPS Sync

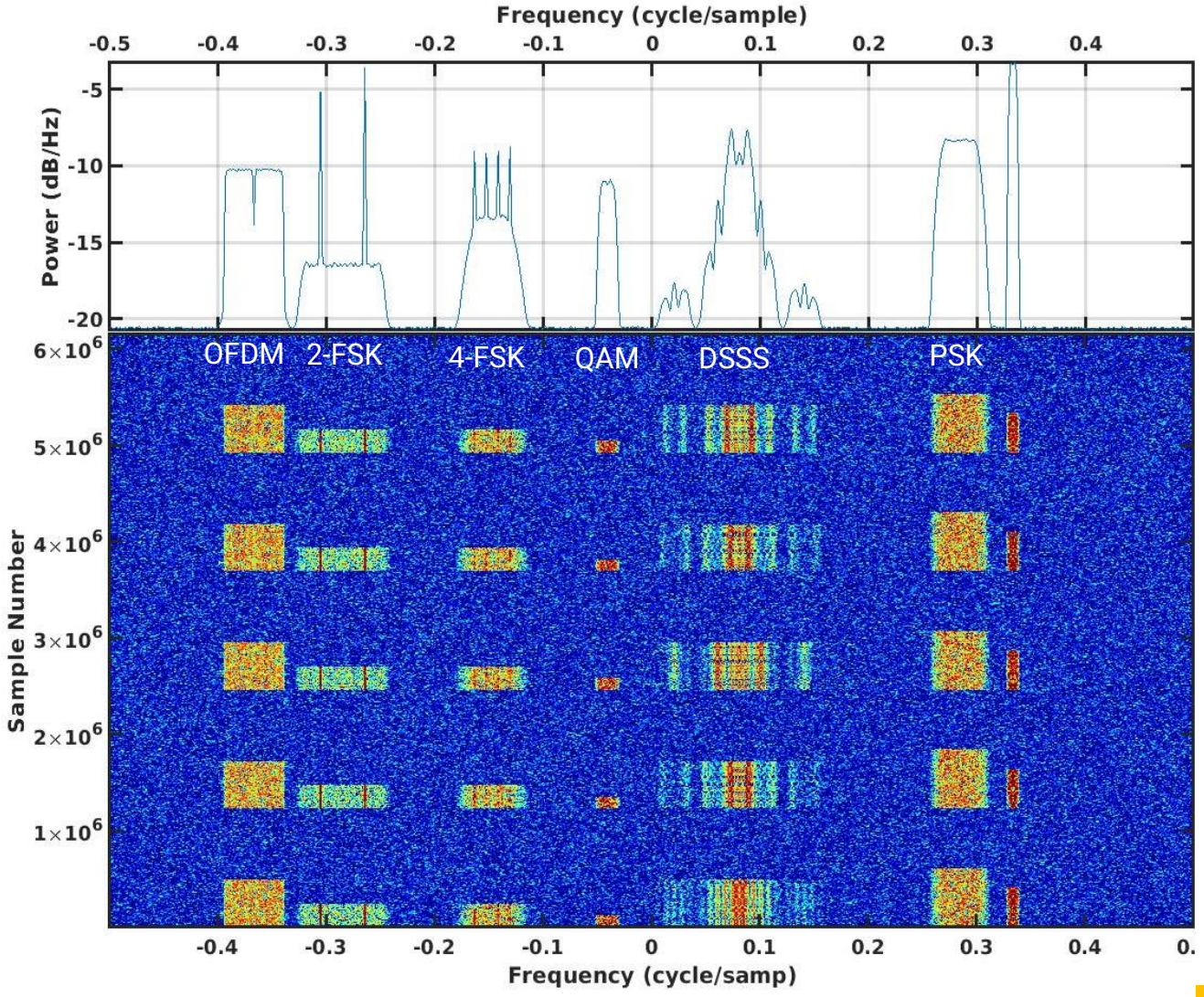


Anechoic RF environment for calibration and measurement

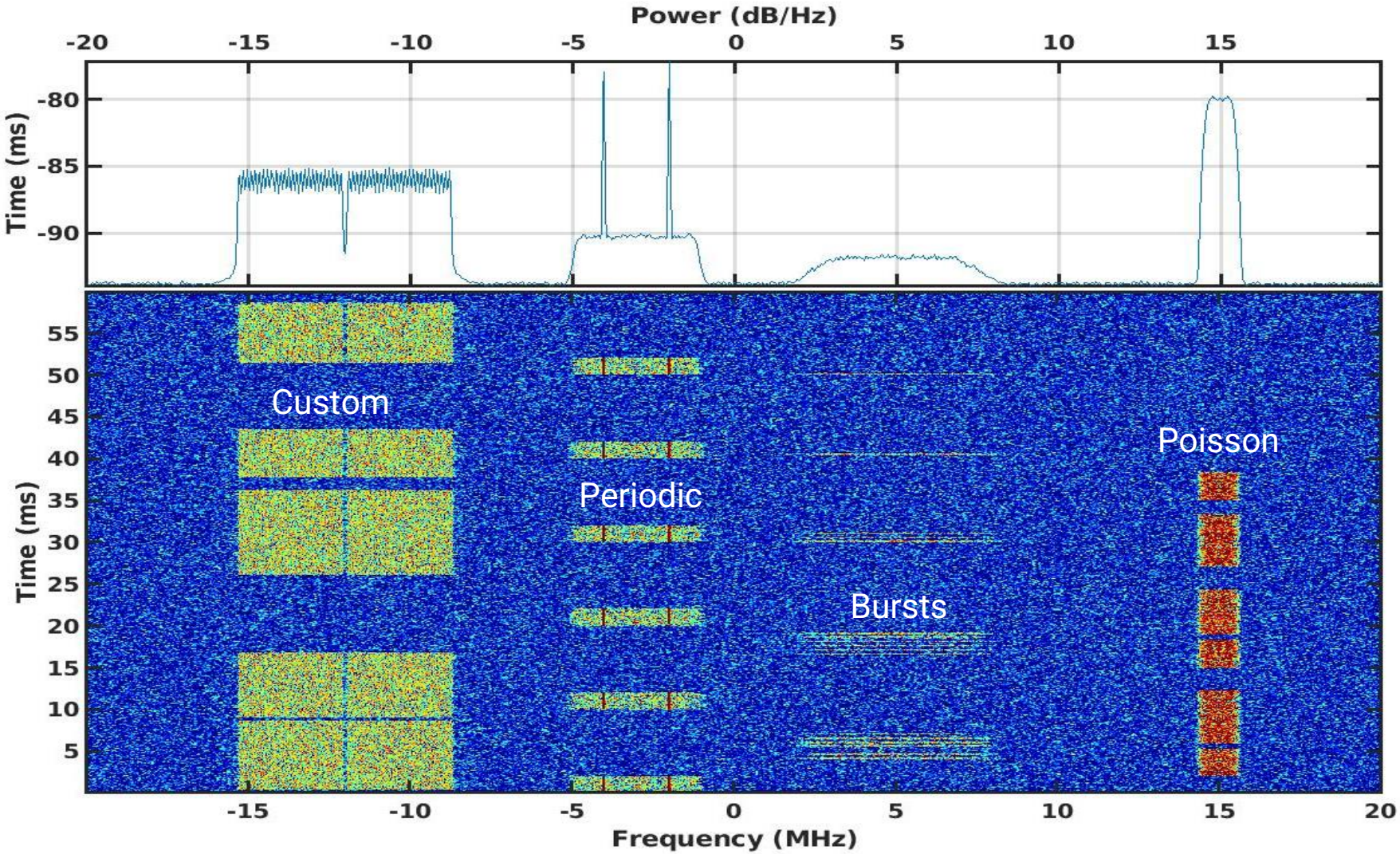


Data generation, visualization, high performance compute

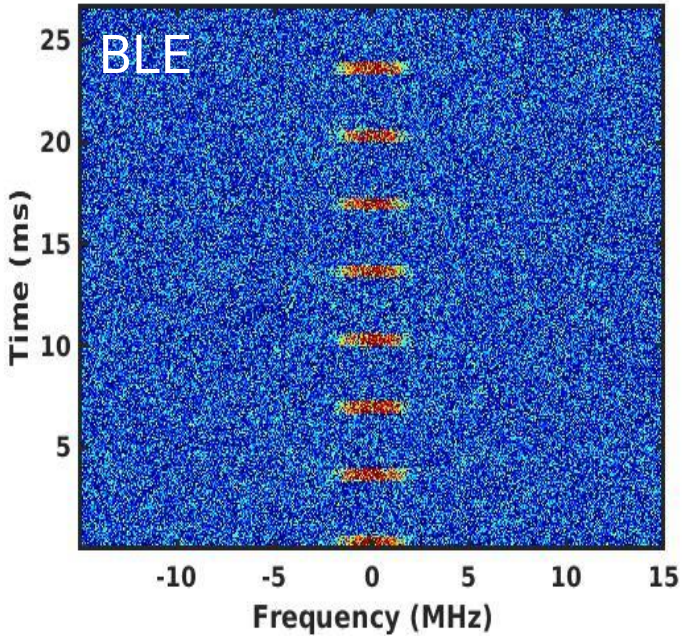
Gallery: Modulation types

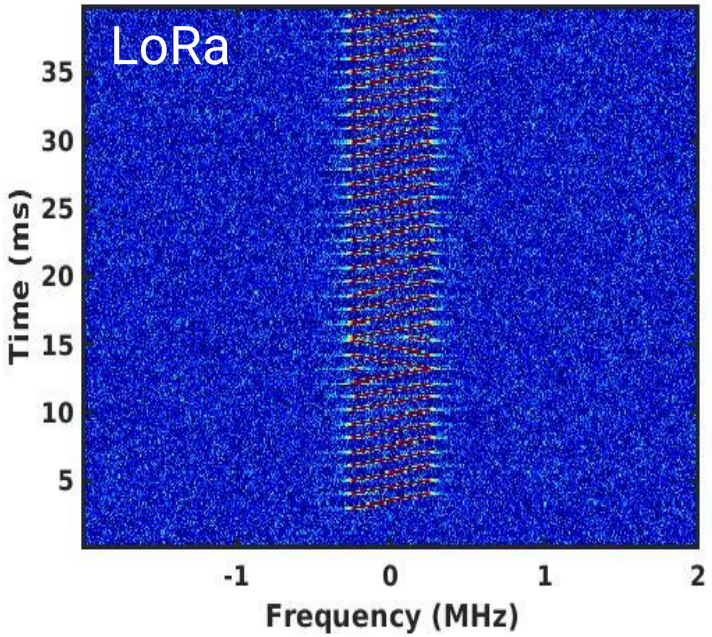


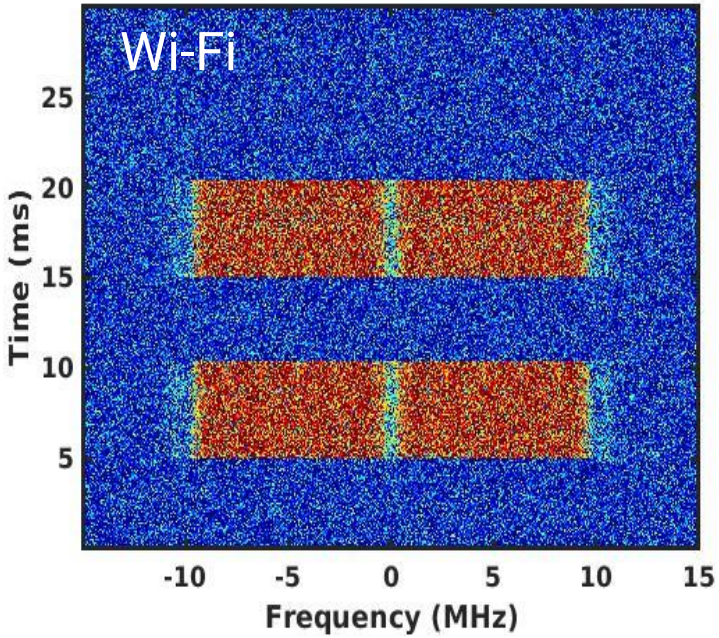
Gallery: traffic generation



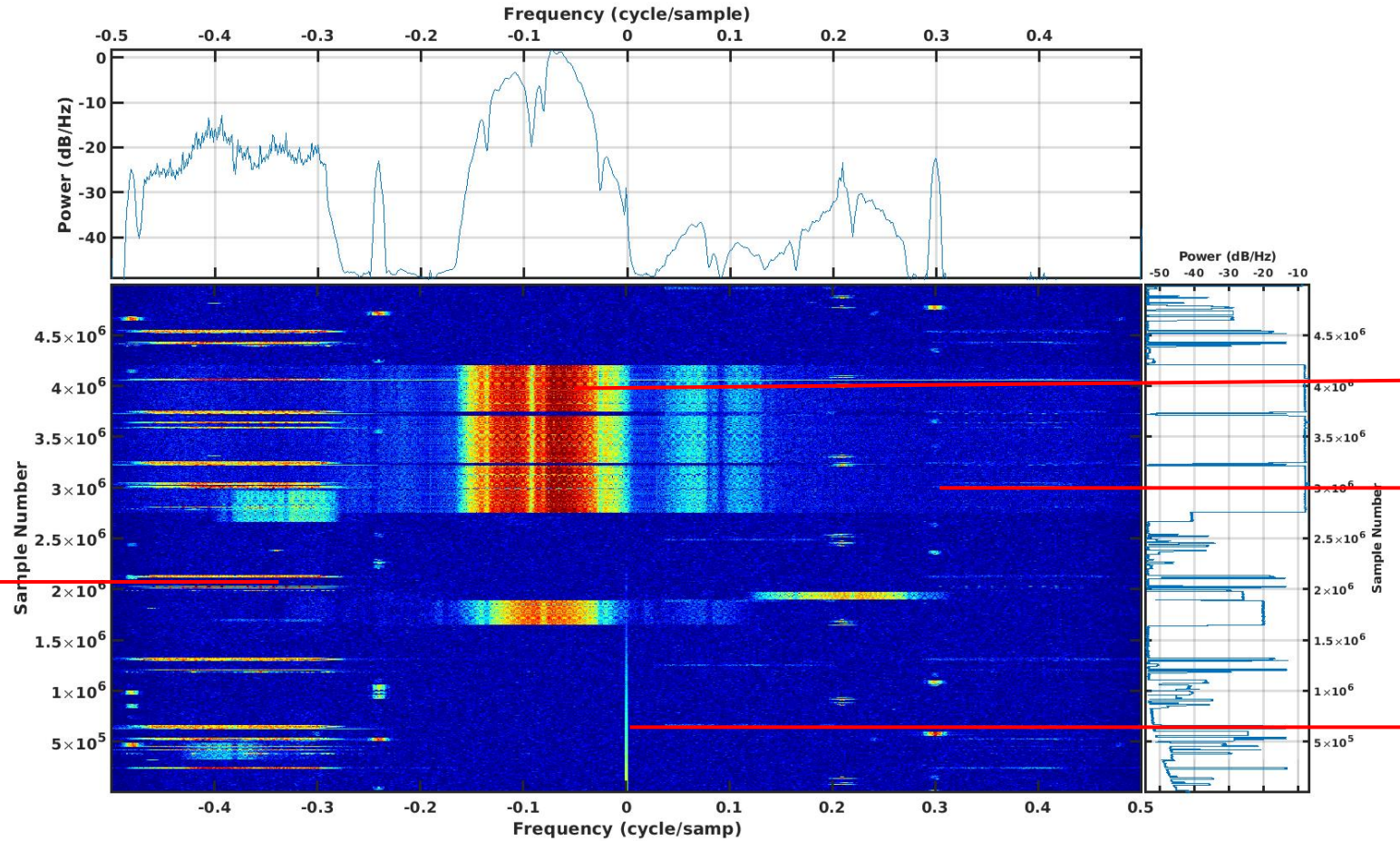
Gallery: protocol compliant signals





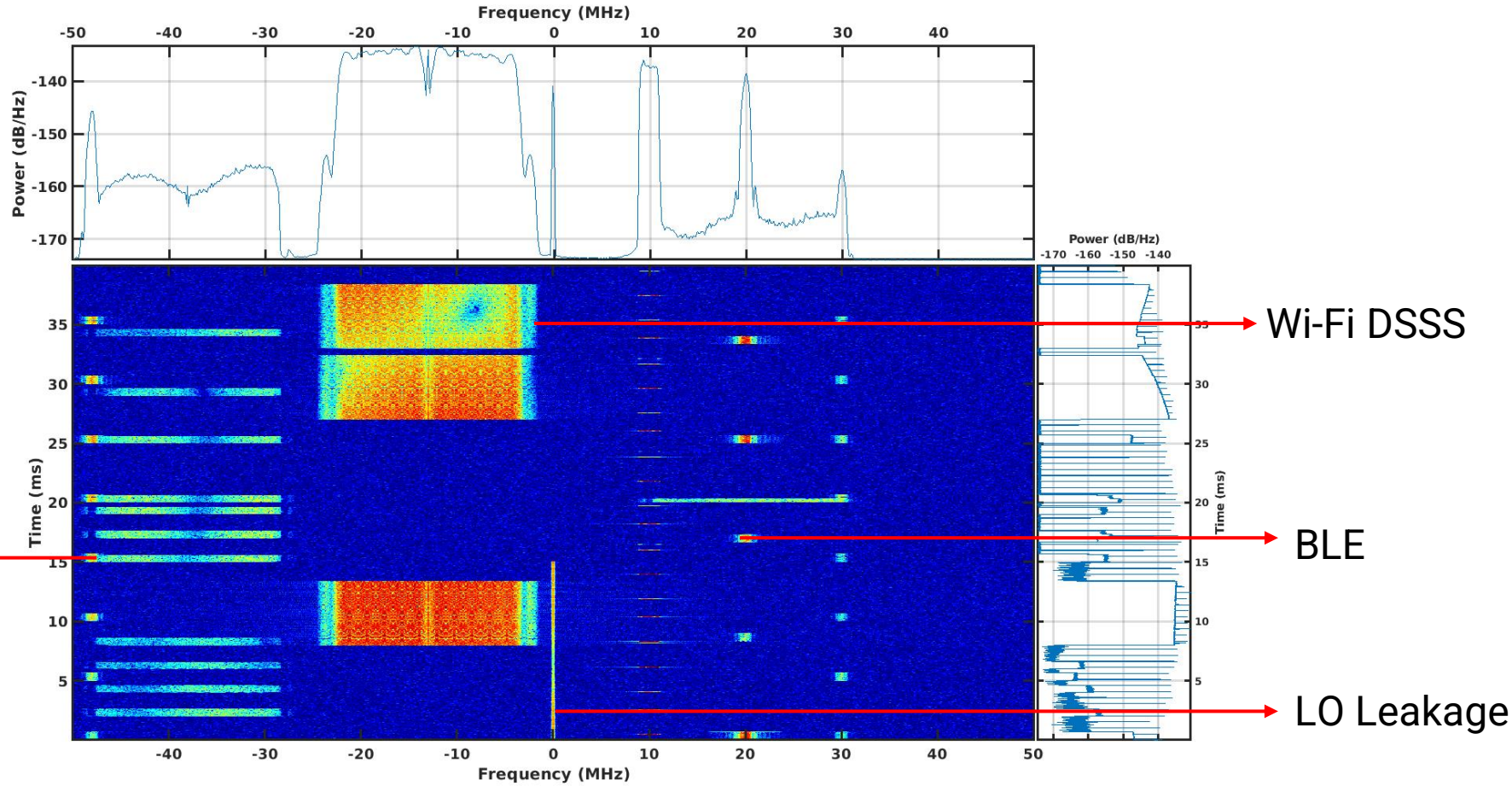


RF band re-creation : 2.45 GHz spectrum collect

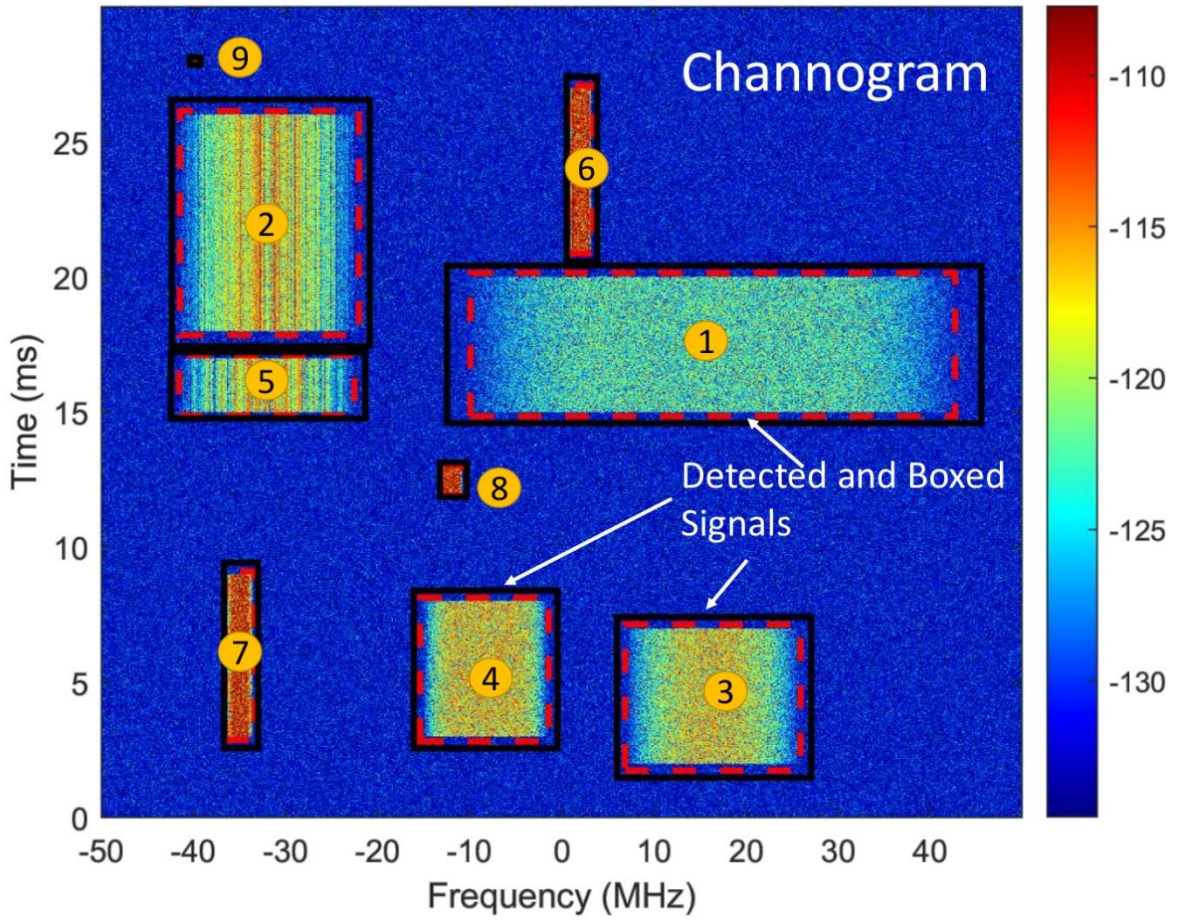


Wi-Fi OFDM

RF band recreation : 2.45 GHz spectrum re-created



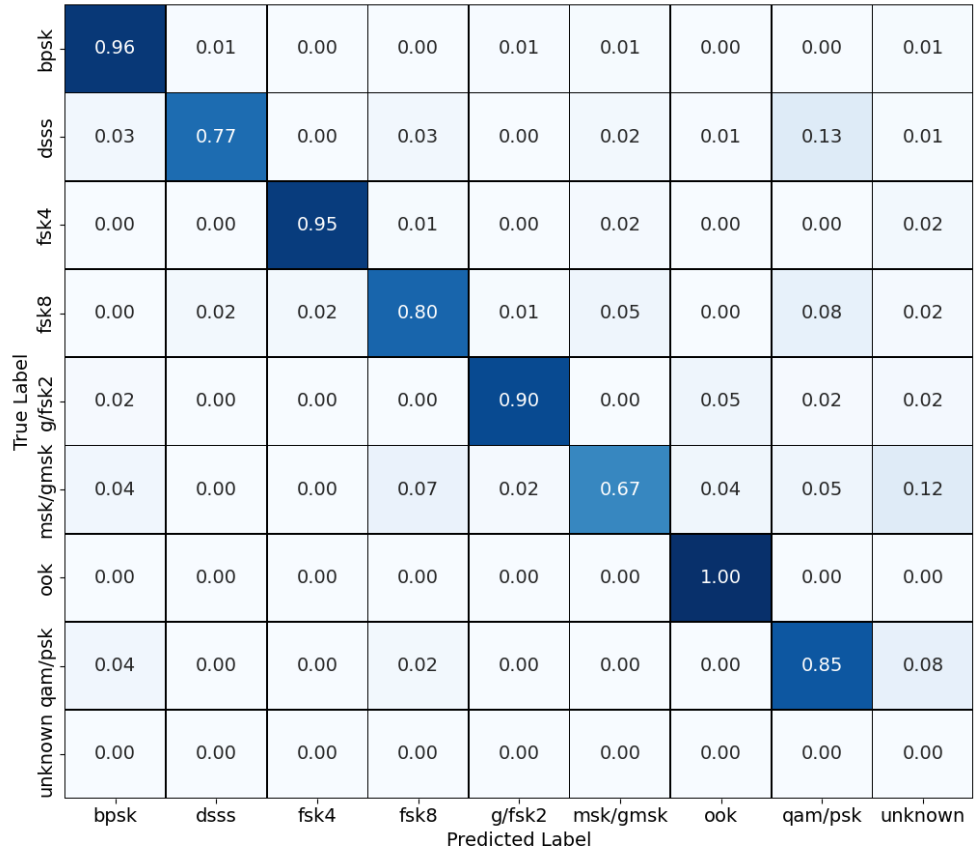
RFSynth for energy detector development



R Bell, K Watson, T Hu, I Poy, F Harris, D Bharadia *Searchlight: An accurate, sensitive, and fast radio frequency energy detection system*. IEEE MILCOM 2023

RFSynth for automatic modulation recognition

Parameter	RML 2016.04C dataset*	RFSynth
Modulation	11	27
Protocols	N/A	LTE, 5GNR, BLE, Zigbee, LoRa, WiFi
SNR points	-20:18 (inconsistent across modulations)	Configurable
Sample Length	128	Configurable

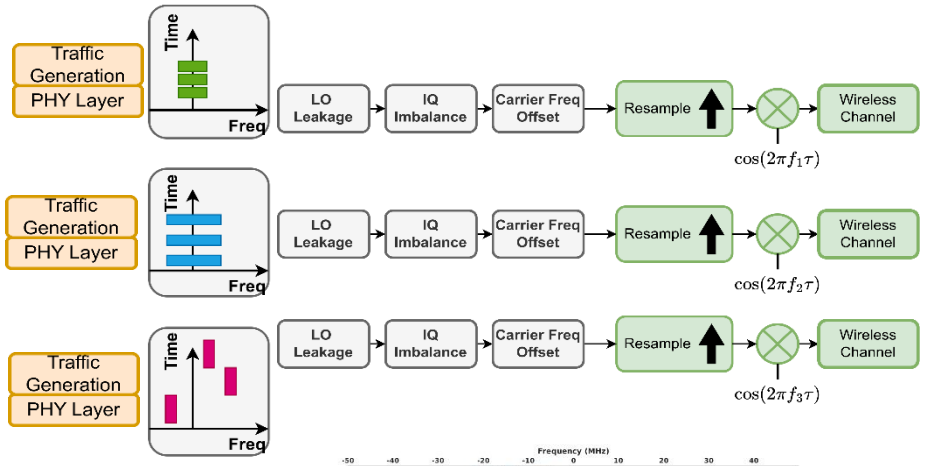


* - T. J. O'Shea, J. Corgan and T. C. Clancy, "Unsupervised representation learning of structured radio communication signals," *2016 First International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE)*, Aalborg, Denmark, 2016, pp. 1-5, doi: 10.1109/SPLIM.2016.7528397.
* - deepsig.ai

R Mathuria, S Rajagopal, D Bharadia, *Fourier Meets Gardner: Robust Blind Waveform Characterization*. IEEE DySPAN 2024

RFSynth: framework for dynamic spectrum data generation

- Generate labelled datasets with multiple simultaneous tx
- Create dynamic spectrum environments and transmit them in your testbed
- Foundation for digital twin and data driven development



Code

github.com/ucsdwcsng/rfsynth

Questions?

Acknowledgement

https://www.iarpa.gov/research-programs/scisrs

This research is based upon work supported in part by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via [2021-2106240007]. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes not withstanding any copyright annotation therein.